Traits and adaptations

Traits and Adaptations:

A theoretical examination and new empirical evidence

Samuel Henry

University of Edinburgh

René Mõttus

University of Edinburgh and University of Tartu

* Department of Psychology, University of Edinburgh

S6, 7 George Square EH8 9JZ Edinburgh, UK +44 745 332 8181 | samuel.henry@ed.ac.uk

** Department of Psychology, University of Edinburgh

S4, 7 George Square EH8 9JZ Edinburgh, UK rene.mottus@ed.ac.uk

Abstract

We investigated the distinction between *traits (*also labeled *basic tendencies* or *dispositions)* and *(characteristic) adaptations,* two related features of the personality system postulated to influence how personality manifests throughout the lifespan. Traits are alleged to be universal, causal, and enduring entities that exist across cultures and through evolutionary time, whereas learned adaptations are acquired through sustained interaction with cultural, physical, and social environments. Although this distinction is central to several personality theories, they provide few measurable criteria to distinguish between traits and adaptations. Moreover, little research has endeavored to operationalize it, let alone test it empirically. Drawing on insights from four frameworks—the Five-Factor Theory, Cybernetic Big Five Theory, Disposition-Adaptation-Environment Model, and New Big Five—we attempted to investigate the distinction both theoretically and empirically. Using various experimental rating conditions, we first scored 240 questionnaire items in their degrees of definitionally reflecting traits and/or adaptations. Next, we correlated these definitional ratings with the items' estimates of rank-order stability, consensual validity, and heritability—criteria often associated with personality traits. We found some evidence that items rated as more trait-like and less adaptation-like correspond to higher cross-rater agreement and stability, but not heritability. These associations survived controlling for items' re-test reliability, social desirability, and variance. The theoretical and empirical implications of these findings are discussed.

Keywords: Trait, Characteristic Adaptation, Five-Factor Theory, Cybernetic Big Five Theory, Duality Principle

Traits and Adaptations: A theoretical examination and new empirical evidence

Several prominent personality theories (e.g., McAdams & Pals, 2006; McCrae & Costa, 2008; DeYoung, 2015; Asendorpf & Motti-Stefanidi, 2018) rest on a distinction between two features of the personality system, which we refer to as *traits (*or *basic tendencies, dispositions)* and *adaptations (*or *characteristic adaptations).* Traits are considered abstract, causal, and enduring entities that exist across cultures and have emerged over evolutionary time in response to universal stimuli. Through sustained interaction with one's particular cultural, physical, and social environment, they become expressed as adaptations—each individual's unique psychological characteristics, ranging from goals and habits to the self-concept[1]. The trait-adaptation distinction mirrors the one between core and surface traits (Cattell, 1946; but see Kandler, Zimmerman, & McAdams, 2014).

At first glance, it provides an elegant way to reconcile observations of numerous highly universal features of personality variability—which point to its deep evolutionary ("biological") roots (Allik, Realo, & McCrae, 2013)—with the inevitable context-specificity of how personality is expressed and measured: the former pertain to traits, the latter to adaptations. At a closer look, however, the distinction becomes less clear, because while theorists do provide lists of examples for each, they offer few well-defined and testable criteria for telling traits and adaptations apart; measurable criteria may exist for traits but less so for adaptations (cf. Asendorpf & Motti-Stefanidi, 2018). The paucity of such criteria complicates exploring the conceptual and practical value of the trait-adaptation distinction. Clarifying the distinction is therefore of utmost theoretical importance.

[1] Theories differ in what their proponents include among adaptations. For example, the Five-Factor Theory (McCrae & Costa, 2008) has a wide range of psychological characteristics categorized as adaptations, while the Cybernetic Big Five Theory (DeYoung, 2015) delineates only three categories of adaptations: goals, interpretations and strategies.

Moreover, recent findings show that hundreds of questionnaire items contain unique variance with properties typically associated with traits. These include 1) stability over more than a dozen years, 2) agreement across assessment methods, and 3) heritability (Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017; Mõttus, Sinick et al., 2019), as well as 4) distinct developmental patterns (Mõttus & Rozgonjuk, in press) and 5) predictive validity (Seeboth & Mõttus, 2018). This has led some theorists to suggest that items represent both traits and adaptations at the same time (Costa & McCrae, 2017). While this may blur the line between them, it also offers an opportunity to operationalize and empirically explore the trait-adaptation distinction: plausibly, items can be quantified in the degrees to which they definitionally correspond to either type of construct, and these degrees can be linked with the empirical properties of the items such as their stability, cross-method agreement, and heritability.

Traits

Five-Factor Theory (FFT; McCrae & Costa, 2008), Cybernetic Big Five Theory (CB5T; DeYoung, 2015) and frameworks such as the New Big Five (NB5; McAdams & Pals, 2006) or the Disposition-Adaptation-Environment Model (DAE; Asendorpf & Motti-Stefanidi, 2018) differentiate latent, biologically-rooted, and humanly-universal personality *traits (or basic tendencies, dispositions)* from their more learned, culture-specific, and circumscribed manifestations—*(characteristic) adaptations (CAs)*.

Personality traits [or basic tendencies (BTs) in FFT and dispositions in DAE and NB5] are commonly defined as latent constructs that dispose an individual to consistent patterns of emotion, motivation, cognition, and behavior (McCrae & Sutin, 2018)—although conceptions of traits do vary within personality research[2]. An empirical hallmark feature of traits is that individual differences in them

[2] Depending on the model, traits can be conceptualized in a variety of ways: emergent from a series of mutually-reinforcing smaller traits, composites of their indicators, descriptive summaries of groups of related behaviors, or even simply 'mathematical fictions' (Revelle & Elleman, 2016). A full consideration of the ways that personality researchers conceive of traits is beyond the scope of this paper, but we

are stable over time (Funder, 1991; McCrae & Costa, 2008). Also, unless there are explicit reasons to believe otherwise, traits should generally be expressed in ways that are comparably detectable by different observational methods such as informant-ratings and self-reports, as otherwise their objective existence cannot be established (e.g., Funder, 1991; Funder, Kolar, & Blackman, 1995). Of course, the criterion of cross-rater agreement suffers from the possible limitation that some personality characteristics may be inherently less visible to external raters than others because of their content (e.g., emotions rather than overt behavior), and this may in principle be confounded with their trait-ness—but we currently have few, if any, alternatives to assessing the method-independence of a wide spectrum of traits.

Additionally, several theorists (e.g., Allport, 1931; McCrae & Costa, 2008) have proposed at least some genetic basis as a necessary cause of variation in traits, consistent with the notion that they are intrinsic to individuals rather than merely learned responses (but for a critique of this criterion see Johnson, Penke, & Spinath, 2011; Turkheimer, Pettersson, & Horn, 2014). Finally, cross-culturally universal properties of traits (e.g., associations with demographic factors such as age and gender) are also consistent with their relative independence from social and temporal context (Allik et al., 2013).

Much of recent personality research has relied on the Five-Factor Model (FFM; McCrae & John, 1992) or the "Big Five" (Goldberg, 1990) domains of Neuroticism (or inversely, Emotional Stability), Extraversion, Openness to Experience (or Intellect), Agreeableness, and Conscientiousness. The FFM domains show substantial rank-order stability over many years (Terracciano, Costa, & McCrae, 2006), cross-rater agreement (McCrae, Costa, et al., 2004), heritability (Briley & Tucker-Drob, 2014), and cross-culturally universal age and gender differences (McCrae & Terracciano, 2005). However, the Big Five

note that distinguishing traits as basic tendencies from their characteristic adaptations generally assumes ascribing traits an underlying causal ontology. However, for a discussion of the ontological status of the 'trait,' see Mõttus (2016) and its subsequent responses.

domains do not exhaust the universe of personality traits. Most FFM traits are at least moderately correlated (van der Linden, te Nijenhuis, & Bakker, 2010) and can therefore be collapsed into broader traits such as stability and plasticity (DeYoung, 2006). Likewise, there are subsets of more strongly correlated characteristics within each of the FFM traits, suggesting that they can be broken apart into narrower traits such as aspects (DeYoung, Quilty, & Peterson, 2007) or facets (Costa & McCrae, 1992). These narrower constructs also demonstrate the same empirical properties associated with higher-order traits (e.g., Jang et al., 1998; McCrae et al., 2005). Such a representation of personality as a system of increasingly narrow traits is called the personality trait hierarchy (Eysenck, 1991).

The trait hierarchy can be extended even below facets. Recent research suggests that specific behavioral, cognitive and affective patterns operationalized with individual questionnaire items such as "I enjoy parties" or "I leave my belongings around" also display stability and cross-method agreement, and most of them also show distinct etiology in terms of heritable variance and disparate developmental patterns (Mõttus, McCrae, Allik, & Realo, 2014; Mõttus, Realo, Allik, Esko, Metspalu, & Johnson, 2015; Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017; Mõttus & Rozgonjuk, in press). This applies even when variance due to higher-order traits (e.g., FFM domains, aspects, and facets) is removed from individual items, suggesting that they contain *unique* information about how individuals differ from each other in stable, observable and heritable ways. These findings as well as items' associations with age and gender tend to replicate across samples from different countries (Mõttus, Sinick, et al., 2019). Items also offer incremental predictive value to a range of life outcomes over the Big Five for both broad and narrow criteria (Seeboth & Mõttus, 2018).

Given that personality is most typically operationalized with questionnaires, items currently constitute the most high-resolution assessments of personality variance and thereby the most specific measurable personality traits. Narrow personality traits currently often captured by single personality test

items have been called *nuances* (McCrae, 2015).

Adaptations

According to the FFT, CAs represent concrete realizations of BTs that emerge via interaction with one's environment across the lifespan (McCrae & Costa, 2008), in a process referred to as accommodation (McCrae & Sutin, 2018). In FFT, CAs include "habits, interests, values, skills, knowledge, beliefs, attitudes, and the internalized aspect of roles and relationships," as well as the self-concept (p. 278; McCrae & Costa, 2008). The DAE model (Asendorpf & Motti-Stefanidi, 2018) describes adaptations similarly to FFT, but postulates that adaptations are mediators between specific traits (dispositions) and specific environmental experiences and can only be defined *relative to them* (i.e., adaptations do not exist independently of what they mediate). Similarly to FFT and DAE, the NB5 model (McAdams & Pals, 2006) describes CAs as "motivational, social-cognitive, and developmental adaptations, contextualized in time, place, and/or social role" (p. 208), although it is less restrictive than FFT and DAE in allowing some CAs to develop independently of traits.

In contrast to FFT and DAE and more similarly to NB5, CB5T (DeYoung, 2015) defines CAs as products of trait-environment interactions that can develop in part independently from their underlying traits—individuals can even adopt goals and strategies that run counter to their traits. However, the CB5T treatment of CAs is more restrictive than those of FFT, DAE and NB5 in that CAs are defined in relation to particular circumstances that are not present for all humans across time (DeYoung, 2015). In other words, any personality characteristic that is culturally unique or would not have existed in human evolutionary history does not meet CB5T's criteria for trait-ness: it is an adaptation. All perspectives but FFT allow bidirectional associations between traits and adaptations; a distinctive feature of FFT is that basic tendencies are insulated from any external influence that does not directly involve altering brain

physiology.

Distinguishing between traits and adaptations

Though several of the discussed theories rest on the distinction, their proponents (e.g., DeYoung, 2015; McCrae & Sutin, 2018; Asendoprf & Motti-Stefanidi, 2018) explicitly recognize that distinguishing between traits and CAs is not straightforward. First, this is because each of the theories defines the constructs slightly differently. Second, this is because the measurements of traits and CAs are confounded: any trait measurement is based on 1) aggregating an individual's adaptations, and/or 2) the individual's assessments of their own traits (e.g., one-item trait scales), which relies on their self-concept that is considered a CA in most theories.

Not surprisingly, then, only few studies to date have directly focused on disambiguating the distinction. Kandler and colleagues (2014) assessed empirical trait criteria such as heritability and stability in what were purportedly either more *core* (e.g., Big Five) or more *surface characteristics* (e.g., attitudes, interests, and goals)—that is, traits and CAs, respectively. They found that some empirical trait criteria (e.g., heritability) were more pronounced in core than surface characteristics, but on the whole, there were few systematic differences between the two kinds of constructs. This led Kandler and colleagues (2014) to call for a revision of the trait-CA distinction.

In defense of FFT, McCrae and Sutin (2018) argued that these results are compatible with the theory:

> *" ... the issue is not whether attitudes, values, and interests are core characteristics, but whether they are [traits] or [adaptations], and the answer to that depends entirely on how one looks at them. If one focuses on the concrete content of a need or interest or value, it is a[n adaptation]; if one focuses on the general tendency underlying the specific*

instantiation, it is likely a [trait].” (p. 160)

This is because according to the *duality principle* of the FFT, any given personality construct (measurement) does two things at the same time: on the one hand, it represents a concrete behavioral (broadly defined) manifestation and on the other, indexes an abstract, underlying trait (Costa & McCrae, 2017). Intelligence offers a useful illustration of this principle. While a score on a vocabulary test indicates learned knowledge due to various environmental factors (years of education, number of books at home, being read to, happenstance), when combined with a variety of other indicators of cognitive skills, it simultaneously serves to indicate an underlying trait of intelligence. [Of note is that some authors object to the notion of underlying intelligence that exists independently of cognitive skills (Kievit, Hofman, & Nation, 2019).]

If only broad aggregate constructs such as FFM domains or even their facets were to meet the empirical criteria for traits, distinguishing them from CAs would be considerably easier. Although individual CAs would be their context-specific indicators, the specifics of individual CAs would be filtered out in aggregation, in line with classical test theory. For example, although a CA itself, the item “I enjoy parties with lots of people” is an indicator of a facet (Gregariousness) and a domain (Extraversion), both defined as the common or compound (McCrae, 2015) variance of numerous indicators that represent these traits. Its CA-ness (reference to parties), thus, would “dissolve” in those higher-order trait scores. However, if the unique variance (above and beyond facets and domains) of such items has properties typically associated with traits, equating traits with higher-order constructs and CAs with the items used to define those constructs is problematic.

According to FFT, the duality principle resolves this problem: every item assesses *both* a trait and a CA, simply depending on how one interprets it (McCrae & Sutin, 2018; see the quote above). FFT does,

however, distinguish between *behavioral* and *global* items. Behavioral items serve as trait indicators, assessing specific activities, interests, goals, and the like, whereas global items refer to broad, adjective-based self-assessments of traits (e.g., "I am extraverted"). Although the latter *seem* to directly refer to traits and bypass CAs, any response is inherently based on one's self-concept and the interpretation of the adjective, and both of these are CAs. FFT refers to this phenomenon as "duality once-removed" (p. 153, McCrae & Sutin, 2018). Regardless of content, each item is equally a trait and a CA, and attempts to distinguish between them are inherently flawed according to FFT.

However, CB5T, NB5, and DAE see constructs as representing either a trait or an adaptation but not both at the same time. This makes exploring their differences a meaningful endeavor according to these theories. For example, DAE offers three specific criteria for separating them: "(i) the adaptation is co-influenced by both the disposition and the environment, (ii) the disposition influences the adaptation more strongly than vice versa, and (iii) the adaptation longitudinally mediates influences of the disposition on the environment or vice versa" (Asendorpf & Motti-Stefanidi, 2018, p. 169). These are useful and clear conceptual criteria for telling the two apart, although empirically distinguishing between them may be more complicated. Because adaptations are only defined relative to specific traits and experiences, and because there is potentially a large number of traits (e.g. nuances) and possibly even an even larger number of environmental experiences linked with these traits, identifying potential adaptations, operationalizing them independently from traits and experiences, and studying their empirical properties in longitudinal designs with appropriately-spaced measurement intervals would be a tall, albeit valuable, order. It may be important to acknowledge, however, that identifying the specific experiences causally linked with personality traits has proven far more difficult than most researchers expected (e.g., Denissen et al., 2019).

The CB5T may offer a simpler solution for operationalizing the difference between traits and CAs: simply classifying constructs as either of them. Items, indexing the narrow traits of nuances, may be

particularly well-suited for this purpose. Their psychological meaning is laid bare without any filters, unlike that of aggregate constructs (e.g., facets or domains) for which item content and verbal label/summary may not entirely match. Also, items are numerous, providing more statistical power to study systematic quantitative variability among them (Mõttus & Rozgonjuk, in press). Indeed, as far as CB5T is concerned, questionnaire items can independently measure either traits or CAs: some items focus "on a particular adaptation, in all its cultural and personal specificity, whereas others 'simply describe culturally universal traits' " (DeYoung, 2015, p. 41). While *variance* due to traits may be part of the variance in CAs empirically—because traits influence the adoption of the latter—the two are still conceptually distinct and can be classified as either by those familiar with their respective definitions.

This study

In an attempt to disambiguate the distinction at least to some extent, the current study aimed to characterize a number of questionnaire items in the degrees to which they reflected a trait and/or an adaptation. We capitalized on previous findings that many individual items reflect specific personality traits with unique empirical properties[3], over and above the broader traits they were designed to measure (Mõttus, Sinick et al., 2019). Every single item can then indicate a BT (a higher-order BT or a narrow BT—nuance—specific to the particular item), a CA (revealing a trait only when aggregated with numerous other adaptations), or both at the same time. Because there are numerous items, we aimed to study systematic variations among them in these properties.

Items were first *definitionally* characterized in their BT-ness or CA-ness by human raters. The

[3] Throughout this paper, we refer to two distinct definitions of "trait": 1) broadly, a feature of a species that displays stability over time, cross-rater agreement, heritability, and cross-cultural validity; and 2) more specifically, one of two distinct categories of relatively stable psychological features utilized by personality theories (distinguished from "adaptations"). When discussing the ideas and findings pertaining specifically to our study, we henceforth use "BT" to denote the latter definition.

raters were assigned to different rating conditions that allowed items to be rated as corresponding to *either* a BT definition or a CA definition, or as simultaneously corresponding to *both* definitions. They were instructed based on BT and adaptation definitions that had been written to be consistent with the relevant major theories and had been discussed with the proponents of the theories[4]. Non-expert ratings were validated against those provided by the authors of the paper and two other researchers who have previously published on the topic. Next, the degrees to which items definitionally reflected BTs and/or CAs (i.e., their ratings) were compared with their rank-order stability, cross-rater agreement, and heritability—properties often associated with BTs (DeYoung, 2015; Kandler et al., 2014; McCrae & Costa, 2008).

We controlled for three potential confounds of the associations: items' social desirability, variance, and re-test reliability. Traits and items are known to vary in their extents of being socially desirable and it is possible that this item property systematically co-varies with other properties, both definitional and empirical; it is conceivable, for example, that traits *per se* are evaluatively more neutral than many of their associated adaptations. Likewise, items with greater variance are likely to have stronger empirical properties (co-variance depends on variance) and they may also vary in the degrees of appearing as traits; for example, social pressure on adaptations may constrain their variance. Finally, more reliably estimable traits may have stronger empirical trait properties and be more likely to appear as traits and/or adaptations definitionally—simply because they are easier to rate.

Our explorations were based on "raw" (as-is) item scores, which contained variance from higher-order traits as well as the nuance-specific variance uniquely associated with each item. Due to complexities associated with interpreting items' unique variance (after having been residualized for facets and domains,

[4] We are grateful to Robert McCrae, Colin DeYoung, Christian Kandler, Jens Asendorpf and Jan-Erik Lönnqvist for their comments on the definitions.

as has been done before for other purposes; e.g., Seeboth & Mõttus, 2018; Mõttus et al., 2017, 2019; Mõttus & Rozgonjuk, in press) in this context, we ultimately chose not to correlate items' BT-ness and CA-ness ratings with the stability, cross-rater agreement, and heritability of their unique variance. If all theories proposed a uni-directional relationship from traits to CAs, we could consider items' residual variance a result of more specific traits, independent of higher-order factors, besides CAs (for an informative overview of parsing item variance, see McCrae, 2015). However, because multiple theories posit a back-channel by which CAs may modify traits, items' residuals become more complicated to interpret (e.g., items could be partly residualized for the possible CA-to-trait effects of their own CA-ness).

Possible predictions

Given that both FFT and CB5T, in particular, *rely* on the distinction between BTs and CAs, we base possible predictions about it primarily on these theories (Table 1). While DAE does specify criteria to examine the distinction, the data we have opted to utilize would not meet the requirements set to distinguish them (see above): it would be difficult to assess environmental influences or longitudinal relationships between these and dispositions/adaptations using a sample of ratings of personality item content.

At face value, different items do appear to assess universally-human and environmentally-adapted characteristics to different degrees (e.g., "I am a worrier" vs. "I love the excitement of roller coasters"), leading us to predict that they would vary in the extents of being rated as corresponding to a BT definition and/or a CA definition, and that different raters and rating conditions could yield roughly similar ratings. According to CB5T (and NB5 and DAE), items rated as more BT-like would likely be rated as less CA-like: a negative correlation between items' CA-ness and BT-ness ratings would be expected. The duality principle of FFT would predict limited variance in the degrees to which items reflect *either* BTs or CAs

and thus no correlation between the ratings of the two, except perhaps for when raters are explicitly asked to contrast them.

By default, one could expect definitionally BT-like items to also display more of the empirical properties typically associated with traits—heritability, stability and cross-method agreement—than the CA-like items. What is definitionally a trait should also be empirically a trait, more so than what is definitionally not a trait. According to both CB5T and FFT, however, the expected pattern may not be as straightforward, and the proponents of both theories have reasons to distance themselves from such specific predictions. In fact, CB5T only explicitly mentions heritability in its section on distinguishing between BTs and CAs, instead preferring the sole criteria of 1) cross-cultural and 2) temporal universality to differentiate them. Regarding genetic influence, CB5T claims that CAs will only appear heritable due to variance they share with traits because, independently of BTs, they are learned responses (DeYoung, 2015). In what complicates things, however, DeYoung (2015) adds:

> "*In CB5T … all of the genetic variance in any characteristic adaptation is a function of related traits, so traits should mediate genetic effects on characteristic adaptations, and one might expect heritabilities to be lower for characteristic adaptations than for traits. In practice, however, it may be exceedingly difficult to identify (and to measure with sufficient accuracy) exactly the relevant set of traits that would account for the genetic variance of any given characteristic adaptation. Heritability studies, therefore, are unlikely to offer any guidance as to what is a trait versus a characteristic adaptation.*" (p. 40)

As neither stability nor cross-rater agreement are explicitly mentioned as criteria to distinguish between BTs and CAs, we derive our CB5T-based predictions of them from personal communications with its author in addition to reading of the theory itself (DeYoung, 2015). Thus, while these predictions

may not (yet) be *explicitly* stated by CB5T, they do reflect the current thinking of its author. Specifically, he postulates that "[CAs] must be stable over time, and given that they are reasonably stable, others are likely to know about them too, yielding inter-rater agreement" and therefore one "would not assume that CAs described in personality items are less heritable, stable, and agreed upon than [BTs], especially in adulthood" (C. G. DeYoung, personal communication, 13 December 2017). In fact, given that CAs are both 1) more specific than BTs and 2) may contrast with their underlying BTs, they may sometimes be even *more* observable to others than traits according to CB5T. While CB5T implies that BTs may be more stable than CAs (simply because BTs—as precursors to CAs—should last longer), its author also notes that "stability is not a criterion that can be used to differentiate [BTs] and CAs" (reviewer comment by C. G. DeYoung, Jan. 2019), as all CAs must be sufficiently stable to be "characteristic" of an individual (DeYoung, 2015).

In summary, therefore, CB5T and its author provide no strong *a priori* reasons to expect correlations between definitional BT-ness on the one hand and stability, cross-rater agreement and heritability on the other. However, since BTs *are* (highly) stable and heritable, whereas only the part of CAs caused by BTs ought to be stable and heritable according to CB5T, stability and heritability could, on average, be smaller in more definitionally CA-like items; we explore these as *ad hoc* possibilities[5].

Likewise, "the most straightforward interpretation of FFT would predict that interests, attitudes, and goals [CAs] ought to be heritable only insofar as they are related to heritable traits" (McCrae & Sutin, 2018, p. 159). That is, definitionally more CA-like items may or may not display heritability (no correlation). However, since FFT BTs are *entirely* heritable, and only the parts of CAs resulting from BTs

[5] Our *ad hoc* predictions are those to which the proponents of the respective theories may not necessarily subscribe to, but that seem worth exploring to us.

are heritable, the average heritability could be expected to be smaller in definitionally CA-like items according to FFT; we consider this as an *ad hoc* possibility (as with CB5T). FFT also holds that "stability is generally expected for [BTs], but it is not contraindicative for [CAs]" (p. 159, McCrae & Sutin, 2018); some CAs may be highly stable (e.g., speaking a language) whereas others (e.g., frequenting a young mothers' club) may be of limited duration. Therefore, although definitionally CA-like items may often be as stable as BT-like items, on average they could be somewhat less stable; again, we explore this as an *ad hoc* possibility (as with CB5T). As FFT offers no specific prediction about cross-rater agreement, we predict no correlation between item's BT- and CA-ratings and cross-rater correlations. Neither *a priori* nor *ad hoc* hypotheses based on any theory were preregistered.

It is noteworthy that while these hypotheses were generated based on extensive consultation with the relevant literature and theorists, they are all essentially null hypotheses—neither theory predicts *any* clear differences between BTs and CAs in their associations with empirical criteria. Specifically, we started the project with stronger hypotheses that seemed plausible at face value; after consulting with theorists and receiving detailed and constructive reviews on the original submission of the article, we only list some possible predictions that we derived from FFT and CB5T *ad hoc*. We thus emphasize that this study is an exploratory investigation of a theoretically very complicated yet extremely fundamental issue in personality science. Our (mostly null) predictions for definitionally and empirically distinguishing between BTs and CAs are preliminary and often *post hoc* to their underlying theories (from consultations with their proponents), but this is exactly the reason for our trying to distinguish them: this complex question has so far largely eluded a clear treatment that entails testable predictions.

[Insert Table 1]

Methods & Participants

Generation of Definitions

We first drafted definitions of BTs and CAs by reviewing relevant literature and consolidating the authors' respective characterizations (i.e., Asendorpf & Motti-Stefanidi, 2018; DeYoung, 2015; Kandler et al., 2014; and McCrae & Costa, 2008). We contacted each of the lead authors, along with one additional expert, to review our descriptions. All experts responded with feedback on the definitions themselves and the project more generally. After revision based on their recommendations, we finalized our definitions as follows:

> **1. Traits:** *Basic tendencies that predispose individuals to their typical thoughts, feelings, and behaviors. It is assumed that traits are innate to all humans, regardless of their particular life circumstances and experiences. People may differ in the magnitude to which they have traits, but the nature of the traits is the same for everyone. You can think of it this way: people from very different cultures may have the same traits, and people many centuries ago also could have had them.*
>
> **2. Adaptations:** *Habitual reactions that people have learned in response to their specific social, cultural, and physical environments. Sometimes, these may be different from what one would predict based on their traits. Unlike traits, people from different cultures and periods of history may have different adaptations.*

BT and CA Ratings of Items

We considered educated but naive-to-the-underlying-theories raters, equipped with a clear task description, suitable judges for rating the degrees to which items' correspond to BT- and CA-definitions: they had limited or cursory prior knowledge on the BT-CA debate and little to no vested interest in it.

However, the items were also rated by a few experts on the topic, allowing us to quantify expert-naive rater agreement. As we expected the task of learning and immediately applying these abstract theoretical concepts to be challenging for raters, we first piloted the task on three individuals to ensure its feasibility (in the *Bipolar* condition—see below). The pilot raters were the two authors of the study and the partner of one of the authors. The partner and the first author had minimal experience with the concepts at the time of completing the task. With a single-rater *ICC* of .44 [95% CI .36, .51] and average-rater *ICC* of .70 [95% CI .63, .76], we judged it likely that there would be enough "signal" in the ratings of multiple raters and chose to proceed with the study. Although we recognize the potential bias involved in the sample used for the pilot, the subsequent relationships between these pilot and ensuing non-pilot ratings (below) lend us more confidence in the task.

We initially recruited 26 individuals to rate the 240 items of the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) in the extent to which they resemble BTs and/or CAs, but the average consensus was generally low—lower than among the ratings in the pilot study. We thus realized a larger participant pool, ultimately recruiting 65 raters. All of the raters had at least some post-secondary education and were aged between 22 and 62 with a median age of 25; 36 had no formal connection to psychology, and all others were either 1) former or current students or 2) otherwise had some experience of psychological research; 2 were experts in having published papers on the BT-CA distinction, and 1 was an expert in being a researcher in psychometrics and personality. Participants were presented with the following text: "Below is a list of items typically used by personality researchers to measure personality characteristics. It is sometimes thought that such items reflect at least two distinct, but related, things." This was followed by the definitions of "traits" and "adaptations" given above.

According to the duality principle in FFT, items assess both BTs and CAs, whereas other perspectives such as CB5T claim that there is a clear distinction between the two. To test these competing

claims, we randomized participants to one of four rating conditions. After reading the definitions, participants received condition-specific instructions which allowed us to assess the extent to which a given item measured BTs *or* CAs (CB5T) or both simultaneously (FFT). In the *Trait* condition (*T; N* = 13), participants rated items as traits on a 5-point Likert scale from 1 ("It does not reflect a trait") to 5 ("It clearly reflects a trait"). The *Adaptation* condition (*A; N* = 13) was identical but asked participants to rate items as CAs. We refer to these conditions as *Independent*. The *Bipolar* condition (*V*; *N* = 24, including pilot raters and expert raters) instructed participants to rate items on a 5-point Likert scale from 1 ("It clearly reflects a trait") to 5 ("It clearly reflects an adaptation"). The *Both* (T_b/A_b; *N* = 15) condition required participants to rate items on both their BT-ness and CA-ness and contained the following instructions after the definitions: "Please rate each item on the extent to which it clearly measures either of them. In addition to measuring clearly either trait or adaptation (use a high score), any item could equally measure both or none of them at the same time (use a low score), or more one than the other (use the appropriate option in the middle of the scale). The scales range from 1 (it does not measure a trait/adaptation) to 5 (it clearly measures a trait/adaptation). For example, if you think an item clearly measures both a trait and an adaptation, you might select "5" for both categories. Similarly, if you think an item somewhat measures a trait but not an adaptation, you might select "4" and "2," respectively." Participants then saw items with options to rate them 1-5 as a trait and as an adaptation. Finally, just prior to being presented the items, participants in all conditions were presented with the following reminder: "Please keep in mind that you are being asked to rate items for their content and not to respond to the item itself." After completing the task, participants were instructed to include their age, gender, and academic/professional background. Given the preliminary nature of the project and the foreseeable difficulty of the task, we also provided participants with a space to leave comments about the study (available in the Online Supplemental Material: https://osf.io/3cebh/). All ratings were completed using

Qualtrics Survey Software with the exception of 2 pilot raters (including the first author) and 2 of the experts, who used a spreadsheet for their convenience.

Heritability, Rank-Order Stability, and Consensual Validity

Items' heritability (based on 2,811 twin pairs), cross-rater agreement (based on 5,421 to 5,441 individuals), and longitudinal stability (based on 1,736 individuals) estimates were obtained from the meta-analysis conducted by Mõttus and colleagues (2019). This study estimated raw and residual variance of NEO-PI-R items in independent samples from Australia, Canada, Czech Republic, Denmark, Estonia, Germany, Japan, and the United States. Full descriptions of the original projects are available in the Online Supplemental Material (https://osf.io/3cebh/).

Reliability

Associations among and between empirical criteria and definitional BT-ness/CA-ness of items could be confounded by their differential reliability (e.g., reflecting differences in ambiguity). Therefore, we aimed to control for inter-item differences in re-test reliability. Mõttus and colleagues (2019) estimated items' retest reliabilities (over one week) to disattenuate their average estimates of items' heritability, stability and consensual validity. However, their estimates were based on a 100-item FFM scale (Goldberg et. al, 2006; Kosinski et al., 2015), so only the average item reliability could be estimated. We, therefore, collected one-week re-test reliability data using Prolific, an online participant-recruitment platform. Our initial sample consisted of 313 individuals. After cleaning the data for un-effortful responding, a sample of N = 263 participants remained (mean age 38.6 years, SD = 11.1; 159 females). All participants worked at least part-time. Participants were re-directed by Prolific to an online form where they completed the NEO-PI-R (Costa & McCrae, 1992) twice within a 7-to-10-day period.

We had reason to believe that parts of our data were susceptible to un-effortful responding. While

administering the survey at Time 1, the survey website suffered from extreme lag due to the high number of participants using the system simultaneously. Thus, we were particularly strict in our assessment of data quality. We used a variety of methods to clean the data for un-effortful responding (e.g., evidence for repeating or random responses), ultimately excluding 50 participants from the original sample. Full details of the data-cleaning process can be found in the Online Supplemental Material (https://osf.io/3cebh/).

In the final data, the median 7-10-day re-test reliability of raw items was .64 (M = .64, SD = .09, range = .36 to .87). This is consistent with estimates that Mõttus and colleagues (2019) reported (a median of .66), which were based on a much larger sample. Given the relatively small sample, standard errors of the correlations in the current study were nearly .05 and 95% confidence intervals therefore nearly .10, but one could expect that the uncertainties of individual estimates tended to cancel out in analyses integrating across the 240 items.

Social Desirability

A well-studied property in which traits and items can differ is their social desirability or evaluativeness (e.g., Allik, Realo, & Mõttus, 2010; Anglim, Morse, de Vries, MacCann, & Marty, 2017). Therefore, we deemed it worthwhile to explore whether the BT-CA distinction in items intersects with their perceived evaluativeness (either desirability or undesirability). We employed the social desirability ratings of the 240 NEO-PI-R items used in Allik and colleagues (2010); these data had been provided by 276 raters from four countries (N = 30, 100, 126 and 20, respectively for Belgium, Czech Republic, Estonia, and Germany) based on four languages (Flemish, Czech, Estonian and German), using a scale from -3 ("extremely undesirable") to 3 ("extremely desirable"). We quantified items' evaluativeness as the average absolute values of their desirability ratings (i.e., deviations from the neutral midpoint, 0), calculated first within each of the four countries and then aggregated into "pan-cultural" average ratings

(i.e., means from each country had equal weights in the aggregate ratings). Although none of the ratings had been based on the English version of the NEO-PI-R, it is plausible that different item translations deviated from English more or less randomly and therefore the pan-cultural evaluativeness scores were comparable to those that could have been obtained based on the English version of the NEO-PI-R.

Variance

We also examined the potentially confounding role of items' variance in associations between item ratings and empirical criteria. Mõttus and colleagues (2019) found that items with more variance had higher meta-analytic estimates for cross-rater agreement $\rho = .52$ ($p < .001$), heritability $\rho = .43$ ($p < .001$), and stability $\rho = .54$ ($p < .001$). However, controlling for items' standard deviations did not substantially affect correlations between heritability, stability, and cross-rater agreement ($\rho = .35$ to $\rho = .43$, $p < .001$), leading them to suggest that "inter-item differences in variance were unlikely to be the only cause for why items systematically differed in their nuancedness" (p. 43, Mõttus et al., 2019). If an item's perceived BT-/CA-ness is one of these causes, then it could be useful to control for items' variance.

Results

All analyses were carried out using R (R Core Team, 2018). All code and data used to produce these results can be accessed in the "Ratings & Analysis" folder in the Online Supplemental Material (https://osf.io/3cebh/); the available files also allow for alternative analyses (e.g., correlating item ratings with residual estimates of empirical properties). We include a spreadsheet ("ItemDataComprehensive.csv") that compiles estimates for rating conditions, evaluativeness, and (raw and residual) estimates of cross-rater agreement, rank-order stability, heritability, re-test reliability, and variance for all 240 NEO-PI-R items. The attached Appendix includes an abbreviated version of this table (also available in the Online

Supplemental Material) containing only estimates of the criteria analyzed in the present study. Descriptive statistics of and intercorrelations among items' ratings, empirical criteria, and potential confounds are included in the Appendix as well.

NEO-PI-R Item Ratings

Table 2 contains estimates of inter-rater reliability (single- and average-rater intra-class correlations; *ICC*) in each condition. Cicchetti (1994) advised that an average-rater *ICC* of less than .40 is poor, .40 to .59 fair, .60 to .74 good, and .75 to 1.00 excellent. According to these criteria, *T, A,* A_b, and T_b ratings had fair average-rater *ICC*s, whereas the consensus for *V (Bipolar)* ratings was excellent. Evidently, rating items in the degrees to which they corresponded to the definitions of BTs and CAs independently was difficult, but raters found it somewhat easier to contrast BTs to CAs than rating any one of them alone.

We then calculated mean item ratings within each condition and correlated these across conditions (Table 3). The average pilot ratings were consistent with the average non-pilot ratings (excluding expert ratings) obtained in the *Bipolar* condition (ρ = .42, p < .001). Items rated as BTs/CAs in the *Independent* condition also tended to be rated as BTs/CAs in the *Both* condition (ρ = .47/.58, p < .001), suggesting convergent validity (Table 3). Conversely, there were negative correlations between mean BT and CA ratings (ρ = -.56 to -.25, p < .001) with the exception of T/A_b (ρ = -.08, p = .224), suggesting that definitionally BT-like items tended to be rated lower in CA-ness. Likewise, the average ratings in the *Bipolar* condition correlated negatively with average BT-ness ratings (ρ = -.48 to -.62, p < .001) and positively (ρ = .46 to .63, p < .001) with average CA-ness ratings in the *Both* and *Independent* conditions.

Expert ratings showed reasonable overlap with both lay and pilot ratings. The ratings of the two experts who have published on the distinction were both in the *Bipolar* condition and correlated ρ = .40

($p < .001$) with one another. The third expert was in the *A* condition; their ratings correlated $\rho = .19$ ($p = .003$) and $\rho = .22$ ($p < .001$) with the others' ratings. We denote the average ratings of the first two E_V, while we label the third E_A. E_V ratings correlated $\rho = .65$ and .47 with average pilot and non-pilot ratings, respectively ($p < .001$). Meanwhile, E_A correlated $\rho = .33$ ($p < .001$) with average *A* ratings. When averaging the ratings of all three experts, these correlated $\rho = -.31$, $\rho = .39$, $\rho = .28$, $\rho = -.40$, and $\rho = .58$ with respective mean ratings for *T, A,* A_b, T_b, and *V* ($p < .001$); all average ratings in these associations were calculated excluding expert and pilot ratings. These results suggest that while nuanced differences in perceptions of the BT-CA distinction may exist between experts and laypersons, expert ratings tend to follow the same patterns as lay ones, and the differences do not appear more pronounced than those between any given group of raters.

Given the substantial *T*-T_b and *A*-A_b associations, we aggregated both BT- and CA-ratings into (*Combined)* T_c *and (Combined)* A_c ratings, respectively. The average-rater *ICC*s for T_c and A_c were *ICC* = .68, 95% CI [.62, .74] and .72, 95% CI [.66, .77]. The T_c and A_c ratings correlated $\rho = -.47$, 95% CI [-.57, -.36] with each other and $\rho = -.65$, 95% CI [-.72, -.56] and $\rho = .60$, 95% CI [.51, .68] with the *Bipolar* ratings ($p < 10^{-10}$). These findings suggest that a) the combined ratings were sufficiently reliable according to Cicchetti's (1994) recommendations for *ICC*s, and b) items rated as corresponding to the definition of a BT were less likely to be rated as corresponding to the definition of a CA. Based on the consistencies found across conditions, we also formed an *overall* CA-ness score for each item (A_O, $M = 0$, $SD = .35$) by combining items' average standardized ratings from each condition (reversing those in the *Trait* conditions). A_O ($N = 80$ partly overlapping raters) had an excellent average-rater *ICC* of .92 (95% CI [.90, .93]) and correlated -.54 and -.81 with *Independent* and *Both Trait* conditions, .77 and .70 with *Independent* and *Both Adaptation* ratings, and .89 with *Bipolar* ratings ($p < .001$).

The five most CA-like items dealt with tolerating controversial ideas, thinking politicians must pay more attention to human needs, thinking that policy should keep up with a changing world, believing religion should not dictate morals, and thinking it is fine for adults to change their mind on what is right. The five most BT-like items were to do with having a bustling imagination, having many emotions, being persistent and unyielding, being full of energy, and loving to talk to people. We also divided items into quintiles based on their A_O rating and examined the distributions of item ratings by domain (Table 4).

Mean ratings for NEO-PI-R domains were A_O = -.08, -.17, -.04, .21, and 0 for Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness, respectively. We conducted a one-way ANOVA across domains and found significant differences in average BT-CA ratings ([$F(4, 235) = 8.58$, $p < 10^{-6}$). Post-hoc comparisons using the Tukey test indicated that the differences between mean A_O ratings for Agreeableness on the one hand, and Extraversion and Neuroticism on the other, were significant after Bonferroni adjustment for multiple comparisons ($p < 10^{-6}$ and $p < .001$, respectively; the adjusted significance threshold was $p = .005$). Thus, while items for Neuroticism and Extraversion were rated as more CA-like and less BT-like than those for Agreeableness, no other inter-domain differences in ratings were significant.

[Insert Tables 2, 3, and 4]

Relationships between Items' Empirical Trait-Criteria and their BT/CA Ratings

We used the T_c, A_c, *Bipolar, and* A_O conditions to examine associations of items' definitional BT-ness and CA-ness with their empirical trait criteria (Table 5).

Items that were rated more CA-like in the combined *Adaptation* condition (A_c) demonstrated significantly lower cross-rater agreement and rank-order stability ($p < .001$). Likewise, items' combined

CA-ness/BT-ness scores (A_O) had negative correlations with agreement and stability (p = .002 and p = .009). No significant relationships were observed in any other conditions.

Items' 7-10-day retest reliability estimates were largely unrelated to their BT-/CA-ratings, correlating ρ = .01 (p = .885), ρ = -.12 (p = .060), ρ = -.04 (p = .538), and ρ = -.06 (p = .348) with T_c, A_c, V, and A_O. However, they had substantial associations with items' empirical properties, correlating ρ = .57 (p = .001), ρ = .56 (p < .001), and ρ = .49 (p < .001) with cross-rater agreement, rank-order stability, and heritability, respectively. After controlling for retest reliability, two further associations (between T_c on the one hand and cross-rater agreement and stability on the other) reached nominal significance (p = .042 and p = .047), forming a mirror-image of the associations with A_c. Otherwise, controlling for re-test reliability had a negligible impact on the associations.

Items' evaluativeness correlated inconsistently with their BT-/CA-ratings, with ρ = .10 (p = .106), ρ = .27 (p < .001), and ρ = -.16 (p = .015) for T_c, A_c and V, respectively; A_O was not related to evaluativeness (ρ = .02, p = .707). However, more evaluative items had lower cross-rater agreement, stability, and heritability—ρ = -.21 (p = .001), ρ = -.27 (p < .001), and ρ = -.19 (p = .003), respectively—suggesting that neutral items tend to convey more consensually valid, stable, and heritable information about people than more evaluative items. Controlling for evaluativeness had no attenuating effect on the associations between either A_c or A_O and empirical trait criteria, but caused the associations between cross-rater agreement and stability on the one hand, and T_c and V on the other, to reach significance (p = .031 and p = .024 for T_c; p = .049 and p = .041 for V).

Items' standard deviations correlated ρ = -.18 (p = .006), ρ = -.19 (p = .003), ρ = .15 (p = .024), and ρ = .03 (p = .640) with T_c, A_c, V, and A_O. Re-calculating associations between items' CA-/BT-ratings and empirical criteria while controlling for items' standard deviations reinforced the pattern of more

adaptation-like items having lower stability and cross-rater agreement: five of eight partial correlations of items' ratings with stability and cross-rater agreement were significant at $p < .001$. Meanwhile, although the association between A_c and stability was attenuated ($\rho = -.16$, $p = .015$), correlations between *V* on one hand, and stability and agreement on the other, reached significance ($\rho = -.20$, $p = .002$; $\rho = -.19$, $p = .003$). The direction of correlation was consistent across conditions (i.e., positive associations with T_c and negative with all others).

Likewise, when simultaneously controlling for re-test reliability, evaluativeness, and standard deviation, seven of eight partial correlations of items' CA-/BT-ratings with cross-rater agreement and rank-order stability were significant at $p < .01$ (the relationship between A_c and stability lost significance: $\rho = -.12$, $p = .060$), with four *p*-values less than .001. *Bipolar* associations with stability and cross-rater agreement were significant at $p = .007$ and $p = .006$, while the correlation between overall adaptation-ness and rank-order stability was significant at $p = .002$.

In summary, items' definitional BT-ness, as opposed to definitional CA-ness, demonstrated significant associations with the empirical trait criteria of higher cross-rater agreement and rank-order stability. The associations were not strong, but they were consistent across different operationalizations of BT-/CA-ness, including when controlling for items' re-test reliability, evaluativeness and/or variance.

[Insert Table 5]

Discussion

Several prominent personality theories rely on the distinction between traits and CAs (e.g., DeYoung, 2015; McCrae & Sutin, 2018; McAdams & Pals, 2006). However, there have only been a few attempts to elucidate specific, empirically testable criteria for distinguishing between the two kinds of

constructs. Moreover, individual questionnaire items that often refer to what many theories would consider CAs at face value (e.g., liking roller coasters) have turned out to capture unique variance (nuances) with empirical trait-like properties that cannot be ascribed to the aggregate personality traits these items were designed to measure, making the trait-adaptation distinction even less self-evident.

Relying on published theoretical views and input from the advocates of several of the relevant theories, we attempted to a) clearly define traits and CAs, b) characterize a number of item-level traits in the degrees to which they corresponded to the definitions of traits (BTs) and/or CAs and c) investigate the associations between the definitional and empirical properties of these item-level traits. Our main findings were that a) items appearing definitionally more BT-like did appear definitionally less CA-like and b) showed somewhat higher cross-rater agreement and long-term stability, but no differences in heritability. Further, c) these relationships were retained—and even marginally strengthened—when controlling for possible confounders such as items' re-test reliability, variance, and evaluativeness. In other words, the more stable over many years and the more similarly observable to people themselves and their informants that narrow personality characteristics—nuances—tended to be, the more they corresponded to what some theorists could consider BTs, or traits.

These findings were not expected *per se* but were consistent across rating conditions, at least when possible confounders were controlled for. To us, the association of definitional trait-ness with stability appears consistent with some of the relevant theories such as FFT and CB5T, even though their proponents have often refrained from explicitly committing to the hypothesis. Perhaps ironically, thus, we may have found more support for these theories than their proponents could hope for.

Methodological considerations

The characterizations of items as corresponding definitionally to BTs and/or CAs were obtained

using ratings of (mostly) naive raters. On the one hand, this may seem like a limitation: why should we trust non-experts for such conceptual judgments? On the other hand, the raters were instructed based on clear and theory-based definitions of BTs and CAs that had been made in consultation with experts beforehand, and we have no reason to assume that the raters *systematically* deviated from these definitions when providing the ratings. Based on the modest inter-rater agreement within most conditions, raters apparently found it hard to judge the degrees to which items reflected BTs and CAs or either of them, although they apparently found it easier to contrast items in these two degrees when directly asked to do so (as evidenced by higher inter-rater agreement in the *Bipolar* condition). However, the plausibly *non-systematic* deviations of the ratings from the "true" degrees of the item properties being rated could cancel out in the average ratings, consistent with a reasonable pattern of convergent and discriminant correlations across rating conditions and acceptable average-rater *ICC*s, especially with data from all conditions combined.

It could even be argued that naïve raters benefited from not being "contaminated" by having worked with the testing material and/or on the particular theoretical question beforehand, and they presumably had limited vested theoretical interest in this research question (unlike experts, possibly)—they just followed the instructions or non-systematically deviated from them. And yet, the independently-provided ratings of the authors of the study converged with the average ratings of the naive raters. Further, ratings provided by experts—researchers who have investigated these theories—did not dramatically differ from those of naïve raters. Overall, thus, we believe that we secured reasonably good estimates of the degrees to which the items were definitionally BTs and/or CAs. Almost certainly, these are the best data currently available on the question.

Could the instructions given to raters and providing them with both BT and CA definitions in all rating conditions have contributed to the BT-ness and CA-ness ratings being negatively correlated? This

seems possible: even in conditions where raters rated only either items' BT-ness or CA-ness, they may have tacitly contrasted the property that they rated to the other property because they had been given definitions of both. This should be considered as a limitation when interpreting the finding that items that appeared more BT-like tended to be less CA-like, and vice versa, although it should not confound our other findings. One way to address this would be to provide two groups of raters with only one of the two definitions, have them rate items on that characteristic (either BT or CA), and compare these ratings both to each other and to our findings.

Hypotheses

Tentatively assuming that the ratings approximated the true values of items' BT-/CA-ness, our results offer some support for the claim made by CB5T (and NB5 and DAE) that personality characteristics represent either a BT *or* a CA, but not both at the same time. Examination of the most extreme items in A_O corroborates this: item content such as beliefs about politics or religion tended to be CA-like, whereas items asking about one's general energy levels or persistence were rated the most BT-like. In line with our definitions, the most BT-like items assess broad, plausibly universal psychological tendencies that any human in any culture throughout evolutionary time could be ascribed to. Meanwhile, CA-like item content is more specific to one's culture, environment, and upbringing.

According to FFT, it could be argued that individuals' responses to even the most BT-like items still partly reflect CAs, representing manifestations of their self-concept through "duality once-removed" (McCrae & Sutin, 2018). A claim consistent with FFT could be that what we measured in the present study was the extent to which items were "global" or "behavioral" (e.g., "I am extraverted" vs. "I talk to lots of people at parties"), rather than quantifying either BTs or CAs. One might test this empirically by asking people to rate items in the degree to which they measure one or the other and then compare these

ratings to our BT/CA-ratings, hypothesizing a positive association between items' BT-ness and globality. However, the extent to which the definitions of item globality vs behavioral-ness meaningfully differ from those of BTs and CAs is not yet clear. Some may even suspect that distinguishing one distinction from another is a semantic question rather than a more broadly theoretical issue.

Although proponents of the main theories that utilize BTs and CAs have commented on how these might correspond to empirical criteria typically used to assess trait-ness, none—both per the theories and based on the kind feedback from the authors themselves—explicitly predicted relationships between items' BT- or CA-ness and any of the empirical trait criteria of heritability, long-term stability, or cross-rater agreement (Table 1). Indeed, the only predictions we made were *ad hoc* to the proponents' theoretical expectations, with explicit reasons for these given in the introduction. Counter to the null hypotheses of both FFT and CB5T, we found associations between items' cross-rater agreement and stability on the one hand and ratings for BT-ness on the other. This was in line with one of our *ad hoc* hypotheses derived from CB5T and FFT that suggested BT-like items may be more stable than those rated as CAs. The positive association between BT-ness and cross-rater agreement was not even hypothesized *ad hoc*. Despite both FFT and CB5T plausibly leading to the *ad hoc* prediction that heritability may be positively associated with BT-ness, our findings failed to support it—no rating condition had significant correlations with heritability. This is consistent with the suggestion that traits' heritability estimates carry little meaningful information about the inherent properties of the traits (Johnson et al., 2011; Turkheimer et al., 2014). Now we have more evidence for this.

Implications for the trait-adaptation distinction

We experienced considerable difficulties with putting forward specific hypotheses regarding the differences between BTs and CAs in their properties, even after having consulted with relevant theorists.

Further, the open comments of participants suggested general confusion over what constituted either a BT or CA. Yet, despite these challenges, a noisy consensus emerged in each rating condition which averaged to an excellent average inter-rater agreement across all ratings. In other words, definitionally distinguishing between BTs and CAs seems possible in principle, even though the judgment may be difficult for any one person for any given construct. Moreover, our findings offer preliminary evidence for the trait-adaptation distinction having some empirical consequences. This should encourage theorists to elaborate on the distinction between traits and (characteristic) adaptations, a recommendation also made by Kandler and colleagues (2014).

Among possible implications, there may be empirically testable consequences to the distinction, if elaborated, clearly defined, and measured—all of which we attempted to do, but which surely can be done better still. For example, it may be that more BT-like constructs have different developmental causes and trajectories than more CA-like constructs. Lumping BT- and CA-like nuances (items) into broad aggregates such as the Big Five may explain why they have been hard to reliably link with specific genetic variants (Lo et al., 2017) or life experience (Denissen et al., 2019) or why these broad aggregates filter out a substantial part of the otherwise measurable age differences in personality (Mõttus & Rozgonjuk, in press). Also, it may be that BT- and CA-like nuances differ in how they are linked with life outcomes in general or with specific types of outcomes; for example, it may be that specific adaptations drive associations with narrow, circumscribed outcomes, while BTs drive associations with broad life outcomes that draw on many trait domains over extended periods of time. In other words, deconstructing and/or differentiating between personality traits according to their BT-ness and CA-ness may open new research avenues into understanding how personality develops and plays out over the life course.

Traits or "Traits"?

One can distinguish between two definitions of the word "trait". First, it can refer to any stable, observable and potentially genetically-based psychological construct. Extraversion is a trait, narcissism is a trait, being securely attached to important others is a trait, supporting conservative politicians and/or pro-environmental policies are traits. Second, a trait can more narrowly refer to BTs as opposed to CAs, with the latter also showing some of the properties typically associated with traits in the first and broader sense.

We have tried to keep these two definitions apart in this paper (e.g., by referring to only the BT-CA distinction when describing our own findings), although we admit to this being very difficult, not least because these two meanings of traits are routinely mixed up in the literature. The empirical nature of the first definition may not be *directly* applicable to the BT-CA distinction; stability, agreement, and heritability cannot be used as criteria to *accurately* distinguish between the two phenomena, even though our findings show that there is some association between stability and agreement on the one hand and BT-ness of traits on the other[6]. Thus, this study did not identify *clear* criteria to distinguish BTs from CAs, but rather investigated whether measurable differences in empirical properties exist between items based on the definitional BT-/CA-ness of their content—and we found some. We would argue that these empirical properties of traits more generally and of BTs in particular—stability and cross-rater agreement—can already offer *some* guidance in navigating the kinds of personality constructs used by psychologists, and should they be conceptually elaborated on in the future, the guidance may become handier still.

If we were to revert to and stay with the original stances of FFT and CB5T in that these empirical

[6] According to one reviewer, "[b]y definition, empirical trait-properties for the first definition of 'trait' are not empirical trait-properties for the second definition."

criteria are of little use, this would present a conundrum for the BT-CA distinction—and for the authors of this paper. Much of the research on the Big Five specifically and the trait hierarchy more generally has focused on heritability, rank-order stability, and cross-rater agreement as evidence for the ontology of traits (e.g., Funder, 1991; McCrae & Costa, 2008). But if these properties cannot, after all, be used to distinguish BTs from CAs, then what can? Or, if there is no point in seeking empirical criteria to distinguish between the two at all (as one reviewer suggests), then what purpose does the distinction serve, other than to be a theoretical carpet under which to sweep otherwise inexplicable patterns?

Theories should be falsifiable, and achieving this begins with delineating empirical criteria with which to ultimately test them. That we encountered such difficulties in generating (largely null) hypotheses that would allow empirically separating traits and adaptations—*including* after personal communications with and reviews from several of the authors of the theories—indicates that the distinction may be too vague in its current form. While we do recognize the conceptual convenience—or indeed necessity—of distinguishing between the two kinds of constructs, we cannot be certain whether the distinction in its current form clarifies or in fact obscures our understanding of the personality system. For example, non-theorists should be able to readily draw predictions from a theory without having to consult its author(s) (even if the authors are extremely forthcoming in their assistance). Thus, we urge theorists to more clearly delineate better-formalized and empirically-testable definitions of the distinction between BTs and CAs.

Narrow Constructs, Broad Content

To the extent that de-contextualized BT-like items may convey somewhat more (empirically) trait-like information, this study seems to offer evidence against the use of contextually specific items and for the use of more global and universal items. Moreover, despite the associations between evaluativeness,

variance, and re-test reliability on the one hand with stability and observability on the other, items' ratings for BT/CA-ness did *not* show systematic relationships with any of the former. In other words, utilizing items that are assessments of more global and universal personality constructs appears to improve empirical validity (in the sense of stability and agreement) beyond the effect of simply being "good" items—stable over short periods, non-evaluative, sufficiently variable.

This finding, while possibly in line with theorists' conceptions of traits, may feel somewhat counter-intuitive, given the recent evidence that *narrower* personality constructs (nuances) contain incremental information over and above higher-order traits and facets. Early evidence suggests that outcome-predictions are mostly driven by items' unique variance (rather than variance shared with other items and thereby defining higher-order constructs; Seeboth & Mõttus, 2018). In this sense, would it not be more logical that specific, contextualized personality traits provide more information than abstract descriptors? Apparently not. Perhaps within the level of nuances, there can exist a useful balance between specificity and breadth. Or perhaps the items' useful unique variance is not their CA-ness but in their unique BT-ness. Future research should investigate whether it is more BT- or CA-like items—if either—that provide better predictions of life outcomes. Indeed, it would seem that anything providing more information—which may or may not often mean BT-like items—will generally have an upper hand.

In what other domains might the content of BT- and CA-like items meaningfully (empirically) differ? We considered three possibilities (evaluativeness, standard deviation, and re-test reliability) and mentioned "globality" vs. "behavioral-ness," but there are a host of other rateable properties that items contain and that we could not examine in the present study. For example, one might also consider the psychological modality—the extent to which an item assesses affect, behavior, cognition, or motivation (desire) (Wilt & Revelle, 2015)—or other features on which items vary such as importance, level of abstraction, state-ness vs. trait-ness, or base rate (e.g., Leising, Scharloth, Lohse, & Wood, 2014), among

others. Future research should examine relationships between such properties and those that we considered here—stability, cross-rater agreement, and heritability—to better understand the unique, myriad ways item content behaves in the personality system. In the long run, this may help with studying the principles of personality development and intersections with life outcomes.

On the usefulness of re-test reliability

A corollary finding of this study was that items' short-term re-test reliability was a fairly strong predictor of other empirical item properties such as cross-rater agreement, long-term stability, and heritability—these were among the strongest associations reported in this study. This, consistent with previous facet-level findings of McCrae and colleagues (2011), may prove to be a very useful finding in that it allows predicting which items convey more information about individual differences than others: items with low-retest reliability are probably of lower quality (e.g., ambiguous or without clear psychological referents for raters) or pertain to traits that are inherently harder to reliably measure with self-reports. (Of note is that Big Five domains explained only 6% of the variability in re-test correlations, whereas facets explained 24%.) Test-constructors are thus encouraged to test for the re-test reliability of individual items and, if possible, not use those with low reliability—they are less likely to carry useful information and could be replaced with those that do convey additional reliable information (McCrae & Mõttus, 2019).

Conclusion

Advancement of knowledge is promoted through informed, open discourse, and theories are written to be tested. Yet, despite its various conceptualizations, the BT-CA distinction has largely eluded empirical scrutiny so far. In this study, we offered an initial investigation with the goal of furthering the

theoretical conversation with novel empirical evidence—to some effect, judging from reviewer comments and our fruitful interactions with the proponents of some of the most relevant theories, for which we are extremely grateful.

According to our results, narrow traits—in the form of items—that raters perceived as measuring something more universal-looking and applicable throughout human evolutionary history tended to be somewhat more 1) stable over long periods and 2) observable to others in ways that were consistent with self-perceptions. This suggests that definitional BT-ness indexes a unique property of personality items that appears to be at least partially linked to the items' empirical properties. The most straightforward explanation for this would be that items asking about broad, universal characteristics are indeed measuring something more stable and observable—a unique personality trait. On the one hand, this may seem unsurprising. On the other hand, given the current murkiness of the CA-BT distinction, this may be nothing short of snatching a victory from the jaws of defeat—we now have at least one empirical handle on the distinction from which to work. Not a perfect handle, but one all the same. Hopefully, our work will help to elucidate this predominantly under-explored question of personality science.

References

Allik, J., Realo, A., & McCrae, R.R. (2013). Universality of the Five-Factor Model of Personality. In P.T. Costa & T. Widiger (Eds.) *Personality Disorders and the Five Factor Model of Personality*. (pp. 61-74). Washington, DC: American Psychological Association.

Allik, J., Realo, A., & Mõttus, R. (2010). How people see others is different from how people see themselves: A replicable pattern across cultures. *Journal of Personality and Social Psychology, 99, 870-882.*

Allport, G. W. (1931). What is a trait of personality? *The Journal of Abnormal and Social Psychology, 25*, 368-372.

Anglim, J., Morse, G., de Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing Job Applicants to Non-applicants Using an Item-level Bifactor Model on the HEXACO Personality Inventory. *European Journal of Personality, 31*, 669-684.

Asendorpf, J. B., & Motti-Stefanidi, F. (2018). Mediated disposition-environment Transactions: The DAE Model. *European Journal of Personality, 32,* 167-185.

Briley, A. D., & Tucker-Drob, E. M. (2014). Genetic and Environmental Continuity in Personality Development: A Meta-Analysis. *Psychological Bulletin, 140*: 1303-1331.

Cattell, R. B. (1946). Personality structure and measurement: I. The operational determination of trait unities. *British Journal of Psychology, 36,* 88-103.

Cicchetti, D. A. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology. *Psychological Assessment 6,* 284-290.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Fl.: Psychological Assessment Resources.

Costa, P. T. Jr., & McCrae, R. R. (2017). The NEO Inventories as instruments of psychological theory. In T. A. Widiger (Ed.), *Oxford handbook of the Five-Factor Model* (pp. 11-37). New York: Oxford University Press.

Denissen, J. J. A., Luhmann, M., Chung, J. M., & Bleidorn, W. (2019). Transactions between life events and personality traits across the adult lifespan. *Journal of Personality and Social Psychology, 116,* 612-633.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138-51.

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33-58.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880-896.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12*, 773-790.

Funder, D. C. (1991). Global traits: A Neo-Allportian approach to personality. *Psychological Science, 2*, 31-39.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656-72.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*, 1556-1565.

Johnson, W., Penke, L., & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality, 25*, 254-266. DOI: 10.1002/per.836

Kandler, C., Zimmerman, J., & McAdams, D. P. (2014). Core and surface characteristics for the description and theory of personality differences and development. *European Journal of*

Personality, 28, 231-243.

Kievit, R. A., Hofman, A. D., & Nation, K. (2019). Mutualistic Coupling Between Vocabulary and Reasoning in Young Children: A Replication and Extension of the Study by Kievit et al. (2017). *Psychological Science,* Advance online publication. doi:10.1177/0956797619841265

Kosinski, M., Matz, S., Gosling, S., Popov, V. & Stillwell, D. (2015). Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. *American Psychologist, 70*, 543-556.

Leising, D., Scharloth, J., Lohse, O., & Wood, D. (2014). What Types of Terms Do People Use When Describing an Individual's Personality? *Psychological Science, 25*, 1787-1794.

Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., … Chen, C.-H. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genetics, 49,* 152-156.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*, 204-217.

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*, 97-112.

McCrae, R. R., & Costa, P. T. (2008). Empirical and theoretical status of the five-factor model of personality traits. In B. Boyle, G. Matthews, & D. Saklofske (Eds.), *The SAGE handbook of*

personality theory and assessment: Volume 1 — Personality theories and models (pp. 273-295). London: SAGE.

McCrae, R. R., Costa, P. T., Jr., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., Hrêbíčková, M., & Urbánek, T. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38*, 179-201.

McCrae, R. R. & John, O. (1992). An introduction to the Five-Factor Model and its Applications. *Journal of Personality, 60*, pp.175-215.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity. *Personality and Social Psychology Review, 15,* 28-50.

McCrae, R. R., & Mõttus, R. (2019). What Personality Scales Measure: A New Psychometrics and Its Implications for Theory and Assessment. *Current Directions in Psychological Science, 28,* 415-420.

McCrae, R. R., & Sutin, A. R. (2018). A Five-Factor Theory perspective on causal analysis. *European Journal of Personality, 32*, 151-166.

McCrae, R. R., & Terraciano, A. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407-25.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project

(2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547-561.

Mõttus, R. (2016). Towards more rigorous personality trait-outcome research. *European Journal of Personality, 30*: 292-303. DOI : 10.1002/per.2041

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, 112*, 474-490.

Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality, 52,* 47-54.

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS One*, *10*, e0119667.

Mõttus, R., & Rozgonjuk, D. (in press). Development is in the details: Age differences in the Big Five domains, facets, and nuances. *Journal of Personality and Social Psychology.*

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability and utility of personality nuances. *Journal of Personality and Social Psychology, 117, e35-e50.*

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Revelle, W., and Elleman, L. G. (2016). Factors are still fictions. *European Journal of Personality, 30*, 324-325.

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate prediction and mapping of life outcomes. *European Journal of Personality, 32,* 186-201.

Terracciano, A., Costa, P. T., Jr., and McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32*, 999-1009.

Turkheimer, E., Pettersson, E., and Horn, E. E. (2014). A Phenotypic Null Hypothesis for the Genetics of Personality. *Annual Review of Psychology, 65*, 515-540.

van der Linden, D., te Nijenhuis, J., & Bakker, A. A. (2010). The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality, 44*, 315-327.

Wilt, J., and Revelle, W. (2015). Affect, Behaviour, Cognition and Desire in the Big Five: An analysis of item content and structure. *European Journal of Personality, 29,* 478- 497.

Table 1. *Predictions of associations between item ratings and empirical trait criteria based on the Five-Factor Theory and Cybernetic Big Five Theory.*

Criterion	*Theory*	
	FFT	**CB5T**
Stability	T = A (an *ad hoc* possibility: T >A)	T = A (an *ad hoc* possibility: T >A)
Heritability	T = A (an *ad hoc* possibility: T >A)	T = A (an *ad hoc* possibility: T >A)
Agreement	T = A	T = A

Note. "T = A" predicts no difference between items' BT-ness or CA-ness in the criterion. "T > A" predicts systematic positive associations between the criterion and items' BT-ness. Because neither the DAE nor NB5 theories make explicit predictions about the heritability, stability, or cross-rater

agreement of traits vs. adaptations (DAE explicitly avoids using them as distinguishing criteria), we did not include these theories in our specific hypotheses for this study.

Table 2. *Single and average random rater intra-class correlation coefficients (ICC) for initial rating conditions as well as combined across all conditions.*

Condition	Single-Rater *ICC*	Average-Rater *ICC*
A: Only adaptation rated (*N* = 13)	.09 [.07, .12]	.57 [.49, .64]
A_b: Adaption when both rated (*N* = 15)	.08 [.06, .11]	.57 [.48, .64]
T: Only trait rated (*N* = 13)	.07 [.05, .09]	.48 [.38, .57]
T_b ; Trait when both rated (*N* = 15)	.09 [.06, .12]	.59 [.50, .66]
V: Trait *vs* adaptation (*Bipolar*; *N* = 24)	.16 [.13, .19]	.82 [.78, .85]
A_O : all ratings (*N* = 80)	.12 [.10, .15]	.92 [.90, .93]

Note. T_b and *T* ratings were reversed and all ratings standardized to create A_O. 95% confidence intervals in brackets. $p < .001$.

Table 3. *Descriptive statistics for each rating condition and Spearman's correlations of ratings across conditions.*

Condition	*M*	*SD*	*T*	*A*	T_b	A_b
T: Only trait rated	3.32	0.47				
A: Only adaptation rated	2.85	0.50	-.25			
T_b ; Trait when both rated	3.44	0.46	.47	-.52		
A_b: Adaption when both rated	3.29	0.48	-.08*	.58	-.56	
V: Trait *vs* adaptation (*Bipolar*)	2.68	0.57	-.48	.63	-.62	.46

Note. M = Mean ratings (rated from 1 to 5). *SD* = standard deviation. Positive scores for *Bipolar* indicate Adaptation-ness. With 1 exception, $p < .001$. *$p = .22$.

Table 4. *Distribution of A_O item ratings by Big-Five domain.*

Quintile	N	E	O	A	C
1	13	15	13	2	5
2	9	16	5	3	15
3	10	7	11	9	11
4	12	4	7	15	10
5	4	6	12	19	7

Note. First quintile represents the items with highest BT ratings (i.e., most negative A_O values).

Table 5. *Spearman's correlations between item ratings and empirical trait-criteria.*

	Raw Correlation			Retest Reliability			Evaluativeness			Standard Deviation			Combined		
	r_{CA}	r_{RO}	h^2	r_{CA}	r_{RO}	h^2	r_{CA}	r_{RO}	h^2	r_{CA}	r_{RO}	h^2	r_{CA}	r_{RO}	h^2
T_c	.11	.11	-.06	.13*	.13*	-.07	.14*	.15*	-.04	.25**	.25***	.02	.24***	.24***	-.01
A_c	-.29***	-.23***	-.11	-.27***	-.20**	-.06	-.25***	-.17**	-.07	-.22**	-.16*	-.03	-.22***	-.12	0
V	-.09	-.08	.04	-.08	-.07	.06	-.13*	-.13*	.01	-.20**	-.19**	-.03	-.18**	-.18**	0
A_o	-.20**	-.17**	-.01	-.20**	-.16*	-.02	-.20**	-.17**	-.01	-.25***	-.22***	-.03	-.24***	-.20**	0

Note. Positive V scores indicate adaptation-ness. r_{CA} = cross-rater agreement; r_{RO} = rank-order stability; h^2 = heritability. Column labels (e.g., "Retest Reliability") indicate that associations were calculated while controlling for the listed criterion. * $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix

Table A1. *Estimates of BT-CA ratings, empirical criteria, and potential confounds for the 240 NEO-PI-R items, organized from most to least adaptation-like.*

		Item Ratings				Empirical Criteria			Confounds		
Code	Indications of items (not items themselves)	T_c	A_c	V	A_O	r_{CA}	r_{RO}	h^2	r_{TT}	*eval*	*var*
O6.3	*Believes religion should not dictate morals*	2.21	4.18	4.54	1.22	0.34	0.47	0.57	0.71	1.08	1.03

Traits and Adaptations

O6.2	*Thinks that policy should keep up with changing world*	2.68	4.5	4.29	1.1	0.15	0.29	0.62	0.54	0.84	1.32
A6.1	*Thinks politicians must pay more attention to human needs*	2.86	4.21	4.5	1.01	0.24	0.31	0.44	0.54	0.84	1.55
O6.1	*Tolerates controversial ideas*	2.14	3.86	4.04	1	0.25	0.36	0.42	0.57	0.98	1.39
O6.7	*Thinks it is fine for adults to change their mind on what is right*	2.18	3.68	4.21	1	0.24	0.31	0.45	0.64	1.04	1.05
C1.2	*Fulfils civic obligations*	2.36	3.75	3.96	0.88	0.26	0.28	0.65	0.56	1.23	1.6
A2.4	*Believes that honesty is the best policy*	2.64	4.18	3.75	0.87	0.29	0.27	0.38	0.53	1.05	1.26
O2.3	*Likes expressive dance*	2.25	3.36	3.67	0.72	0.47	0.56	0.67	0.71	1.33	0.7
O6.4	*Is comfortable with people in other societies having different moral principles*	2.86	4.04	3.75	0.72	0.16	0.32	0.33	0.51	0.88	1.61
A6.4	*Feels for beggars*	2.39	3.61	3.79	0.72	0.33	0.4	0.49	0.69	1.03	1.44
O6.8	*Is liberal in moral principles*	2.57	3.57	3.92	0.7	0.26	0.41	0.61	0.58	0.9	0.72
A1.3	*Doesn't think people are after each other*	2.61	3.79	3.5	0.66	0.25	0.43	0.45	0.72	1.06	1.32
A6.3	*Thinks that more should be done for those in need*	2.79	3.57	3.67	0.61	0.23	0.45	0.53	0.61	1	1.23
A6.5	*Puts human consideration ahead of economic benefits*	3.32	3.75	4.08	0.6	0.16	0.37	0.42	0.6	0.88	1.43
E5.2	*Likes garish destinations*	2.93	3.46	3.83	0.58	0.4	0.49	0.61	0.76	1.38	1.13
C3.3	*Doesn't like to owe*	3.14	3.93	3.29	0.53	0.38	0.41	0.58	0.85	1.01	1.98
C3.4	*Doesn't cheat in games*	2.54	3.32	3.46	0.5	0.25	0.32	0.57	0.66	1.16	1.09

Traits and Adaptations

A1.5	*Doesn't worry that kinds acts have ulterior meanings*	3	3.89	3.25	0.5	0.21	0.28	0.41	0.73	0.89	1.26
E4.7	*Lives a fast-moving life*	2.79	3.61	3.42	0.49	0.44	0.44	0.56	0.74	1.1	0.95
A1.1	*Trusts others' intentions*	2.68	3.86	2.83	0.47	0.21	0.41	0.5	0.6	1.04	1.44
O4.4	*Tries different foods*	2.75	3.46	3.21	0.47	0.44	0.52	0.63	0.76	1.2	1.11
C1.6	*Appears successful across the board*	2.46	3.5	3.04	0.46	0.31	0.32	0.5	0.69	1.02	1.2
A6.8	*Would be lenient rather than fair*	2.82	3.54	3.38	0.46	0.21	0.33	0.5	0.46	0.92	0.94
N4.6	*Feels uneasy around authority figures*	3.14	3.71	3.38	0.45	0.32	0.35	0.48	0.75	0.98	1.15
C1.3	*Makes informed decisions*	3.07	3.82	3.04	0.44	0.22	0.35	0.38	0.46	0.77	1.55
N4.1	*Fears embarrassing himself/herself*	3.18	3.61	3.46	0.42	0.23	0.43	0.55	0.65	1.14	1.18
A4.2	*Avoids sarcasm*	3.25	3.89	3.25	0.42	0.33	0.54	0.52	0.73	1.13	1.25
A4.5	*Is forgiving*	3.36	4	3.25	0.41	0.26	0.37	0.44	0.67	1.06	1.32
A3.2	*Is considerate to others*	3.32	3.89	3	0.4	0.25	0.27	0.45	0.61	0.83	1.8
E4.1	*Has an energetic life-style*	3	3.29	3.54	0.38	0.37	0.43	0.54	0.58	1.09	1.06
A4.3	*Avoids expressing anger*	3.11	3.96	2.92	0.37	0.2	0.44	0.48	0.46	1.18	1.11
O6.6	*Is tolerant of different life-styles*	3.36	3.96	2.92	0.37	0.24	0.3	0.55	0.53	0.82	1.96
E6.7	*Uses extremely positive words to describe things*	2.75	3	3.5	0.36	0.24	0.39	0.41	0.46	1.17	0.69
O6.5	*Values sticking with principles rather than being flexible-minded*	3.14	3.39	3.38	0.34	0.23	0.4	0.49	0.56	1	1.06
E2.6	*Likes vacations with crowds*	3.25	3.43	3.63	0.34	0.47	0.52	0.5	0.82	1.29	0.69

Traits and Adaptations

E5.4	*Likes distressing films*	2.86	3.18	3.42	0.33	0.49	0.54	0.65	0.78	1.4	0.73
A5.6	*Feels he/she is not superior than anyone*	3.04	3.68	2.79	0.32	0.2	0.36	0.46	0.52	1.06	1.33
A2.7	*Doesn't bully or flatter to get his/her way*	3.11	3.5	3	0.32	0.24	0.41	0.48	0.7	1.05	1.56
A1.8	*Believes that human nature is good*	2.89	3.5	2.88	0.32	0.25	0.41	0.61	0.67	0.84	1.37
A5.1	*Avoids boasting*	3.29	3.46	3.21	0.31	0.28	0.45	0.49	0.63	1.08	1.43
O1.6	*Doesn't avoid daydreaming*	2.82	2.82	3.42	0.3	0.29	0.36	0	0.59	0.96	0.92
N6.1	*Needs a lot of help*	2.89	3.43	2.96	0.29	0.31	0.33	0.53	0.64	1.03	1.68
C6.1	*Has avoided being reckless*	2.93	3	3.33	0.29	0.3	0.45	0.57	0.75	1.14	0.95
A1.4	*Believes most people can be trusted*	3.18	3.43	3.21	0.29	0.23	0.24	0.56	0.66	0.82	1.38
O2.1	*Values aesthetics*	2.93	3.21	3.08	0.28	0.45	0.43	0.66	0.71	1.21	1.17
C5.1	*Is good at time-management*	3.36	3.46	3.17	0.28	0.34	0.47	0.59	0.69	1.07	1.83
E3.4	*Does the talking in meetings*	3.54	3.54	3.38	0.26	0.4	0.51	0.63	0.76	1.11	0.89
A6.6	*Believes everyone must be treated respectfully*	3.36	3.75	2.83	0.26	0.23	0.38	0.46	0.59	0.88	2.12
A3.7	*Considers himself/herself bighearted*	3.14	3.82	2.54	0.26	0.22	0.31	0.53	0.73	0.81	1.42
N5.1	*Has too much of a good thing*	3	3.18	3.04	0.25	0.22	0.35	0.49	0.51	1.05	0.85
C3.8	*Shows up for work in almost any circumstances*	3.46	3.21	3.38	0.25	0.4	0.39	0.54	0.76	1.11	1.5
E6.1	*Is joyful*	3.11	3.11	3.21	0.24	0.32	0.45	0.61	0.74	1.18	1.24
N4.3	*Has sometimes felt unbearably ashamed*	3.11	3.36	2.96	0.24	0.22	0.42	0.64	0.68	1.2	0.95
N3.4	*Has sometimes felt very guilty and*	2.96	3.36	2.96	0.24	0.32	0.39	0.54	0.64	1.15	1.12

	disgraceful										
C3.6	*Sticks with his/her ethical principles no matter what*	3.5	3.79	2.79	0.23	0.23	0.3	0.53	0.54	0.87	1.18
E4.6	*Is usually in a rush*	2.82	3.21	2.88	0.23	0.32	0.43	0.54	0.62	1.06	1
C2.1	*Prefers to have a detailed plan*	3.36	3.29	3.25	0.22	0.4	0.32	0.44	0.6	1.09	0.9
N3.2	*Feels useless*	2.96	3.32	2.75	0.22	0.32	0.43	0.58	0.8	1.22	1.55
O5.2	*Finds philosophy interesting*	2.89	2.93	3.08	0.2	0.43	0.5	0.65	0.74	1.26	0.92
A5.4	*Tries to be meek*	3.18	3.64	2.63	0.2	0.23	0.42	0.6	0.58	0.9	1.34
C5.7	*Completes even the smallest of tasks*	3.21	3	3.21	0.2	0.25	0.38	0.5	0.57	1.08	1.18
A2.6	*Avoids tricking people into things*	3.21	3.36	2.79	0.18	0.28	0.41	0.53	0.74	1.06	1.39
N4.7	*Finds it very hard to face the person he/she has somehow offended*	3.18	3.32	2.96	0.18	0.17	0.3	0.51	0.62	1.01	1.01
A2.2	*Doesn't manipulate others*	3.21	3.29	2.92	0.17	0.28	0.46	0.55	0.64	1.12	1.79
N3.6	*Has low self-esteem*	3.11	3.39	2.67	0.17	0.32	0.42	0.59	0.75	1.04	1.64
O4.6	*Likes redecorating*	3.11	3.07	3.33	0.17	0.31	0.48	0.6	0.71	1.13	1.02
A3.1	*Is seen as selfless*	3.29	3.46	2.88	0.16	0.27	0.38	0.5	0.7	1.1	1.59
N2.5	*Is often sickened by others*	3.11	3.25	2.92	0.16	0.26	0.34	0.52	0.62	0.96	1.43
A2.8	*Doesn't like to scheme with people*	3.04	3.14	2.88	0.16	0.29	0.41	0.56	0.72	0.92	1.42
E3.2	*Manages to assert himself/herself*	3.64	3.43	3.25	0.15	0.26	0.47	0.49	0.64	1.02	1.19
C3.2	*Is dependable*	3.07	3.11	2.92	0.15	0.25	0.39	0.51	0.66	1.16	1.33
A5.2	*Avoids talking about his/her qualities*	3.21	3.07	2.96	0.15	0.27	0.39	0.51	0.57	1.04	0.86

Traits and Adaptations

N6.7	*Has poor judgement in difficult situations*	3.36	3.21	3.13	0.15	0.21	0.28	0.4	0.58	0.77	1.62
O5.8	*Has curiosity about many things*	3.54	3.39	3	0.15	0.35	0.45	0.56	0.71	1	1.31
O1.2	*Lets imagination fly*	3.46	3.21	3.17	0.14	0.3	0.4	0.63	0.63	1.1	0.98
A2.5	*Wants to be seen true to his/her beliefs*	3.57	3.32	3.04	0.14	0.13	0.19	0	0.63	0.82	1.58
A4.8	*Is rarely part of squabbles*	3.18	3.25	2.79	0.14	0.31	0.34	0.52	0.63	0.98	1.31
A5.3	*Doesn't see themselves as better than others*	3.25	3.29	2.83	0.13	0.21	0.44	0.59	0.71	1.03	1.65
N1.5	*Is concerned about future*	3.11	3.21	2.79	0.13	0.31	0.32	0	0.54	1.08	1.01
O4.2	*Is interested in new hobbies*	3.21	3.25	2.79	0.12	0.36	0.31	0.53	0.53	0.86	1.53
A5.5	*Doesn't have a high self-opinion*	3.14	3.54	2.33	0.12	0.31	0.45	0.57	0.73	1.01	1.41
A1.7	*Assumes people are good*	3.29	3.46	2.58	0.12	0.24	0.29	0.48	0.73	0.85	1.26
C1.7	*Considers himself/herself able*	3.5	3.43	2.83	0.12	0.24	0.37	0.53	0.63	0.8	1.39
A5.8	*Doesn't consider others inferior*	2.82	3	2.54	0.12	0.28	0.46	0.63	0.74	1	1.46
C5.2	*Starts to work right away*	3.25	3.32	2.83	0.11	0.37	0.45	0.66	0.74	1.17	1.44
O5.6	*Is interested in deep questions*	2.93	2.89	2.83	0.09	0.32	0.41	0.57	0.57	1.18	0.95
O2.7	*Can be very aroused by poetry or art*	3.14	3	2.96	0.09	0.4	0.51	0.66	0.82	1.19	0.87
A5.7	*Prefers to recognize others' achievements rather than his/her own*	3.25	3.25	2.54	0.09	0.24	0.42	0.53	0.63	0.94	1.26
C4.8	*Works excessively*	3.5	3.46	2.58	0.09	0.42	0.49	0.56	0.75	1.13	1.24
C3.1	*Handles tasks diligently*	3.61	3.46	2.92	0.08	0.26	0.32	0.48	0.51	0.79	2.03
N4.2	*Feels sheepish around people*	3.29	3.29	2.67	0.08	0.31	0.4	0.54	0.58	1.08	1.19

Traits and Adaptations

A3.5	*Is generous*	3.25	3.25	2.58	0.08	0.19	0.34	0.61	0.55	0.93	1.16
C5.6	*Carries on even when faces challenges*	3.43	3.11	3.08	0.08	0.22	0.32	0.49	0.54	0.98	1.28
N2.7	*Is sometimes spiteful and acrimonious*	3.07	3.14	2.71	0.08	0.27	0.37	0.52	0.63	1.03	1.04
C3.7	*Works carefully to avoid having to do it again*	3.71	3.14	3.17	0.08	0.25	0.33	0.41	0.55	0.79	1.72
O5.4	*Enjoys theoretical conversations*	3.11	2.71	3.04	0.07	0.35	0.49	0.59	0.72	1.12	0.73
O2.8	*Likes emotional and figurative poetry*	3.11	2.54	3.25	0.07	0.35	0.4	0.58	0.68	1.09	0.67
N3.8	*Gets easily disheartened and gives up*	3.25	3.29	2.92	0.07	0.29	0.33	0.51	0.58	1.01	1.47
N4.8	*Feels embarrassment for others*	3.36	2.96	3.17	0.07	0.31	0.32	0.54	0.52	1.03	0.81
O1.4	*Daydreams*	3.25	3.18	2.67	0.06	0.26	0.43	0.53	0.6	1.09	0.89
O2.5	*Is touched by poetry*	3.21	2.68	3.13	0.06	0.45	0.53	0.65	0.72	1.26	0.81
N3.5	*Blames his/herself*	3.36	3.54	2.38	0.06	0.25	0.41	0.42	0.64	1.04	0.94
O4.7	*Vacations in different places*	3.54	2.93	3.17	0.06	0.38	0.45	0.49	0.71	1.08	0.74
E3.5	*Is expected to take lead*	3.54	3.29	2.71	0.05	0.33	0.47	0.46	0.62	0.94	0.85
O2.6	*Finds music fascinating*	3.68	3.04	3.08	0.05	0.38	0.43	0.65	0.78	1.07	1.03
N5.6	*Eats excessively*	2.79	2.79	2.5	0.05	0.37	0.43	0.67	0.72	1.19	1.37
N6.6	*Is unable to self-manage in a crisis*	3.5	3.54	2.58	0.05	0.23	0.32	0.47	0.62	0.83	1.68
C4.2	*Works towards one's goals*	3.68	3.43	2.67	0.04	0.31	0.38	0.54	0.6	1.01	1.64
C1.5	*Is convinced to have a good judgement*	3.25	3.11	2.5	0.04	0.26	0.37	0.46	0.58	0.85	1.2
C2.8	*Doesn't misplace things*	2.93	2.54	2.96	0.04	0.39	0.45	0.53	0.68	1.12	1.19
A1.2	*Believes people have mostly good*	3.32	3.07	2.71	0.03	0.23	0.23	0.43	0.77	0.86	1.45

	intentions										
A4.4	*Doesn't show when dislikes someone*	3.36	3.39	2.46	0.03	0.3	0.36	0.51	0.75	1.05	1.13
C6.8	*Thinks through his/her answers*	3.82	3.46	2.75	0.02	0.2	0.34	0.41	0.49	0.96	1.24
O3.6	*Is emotionally sensitive to environments*	3.25	3.07	2.79	0.01	0.22	0.32	0.58	0.55	0.93	1.03
O3.4	*Monitors his/her feelings*	2.96	2.89	2.5	-0.01	0.27	0.28	0.46	0.61	0.97	1.04
N4.4	*Feels very embarrassed when teased or made fun of*	3.57	3.32	2.5	-0.02	0.25	0.43	0.53	0.61	1.15	1.13
C5.3	*Always finishes the job*	3.61	3.54	2.25	-0.03	0.3	0.45	0.54	0.68	0.83	1.54
N4.5	*Thinks others are better than him/her*	3.18	3	2.33	-0.03	0.33	0.44	0.58	0.68	1.09	1.28
N3.7	*Feels downhearted and futile*	3.21	3.14	2.54	-0.03	0.35	0.4	0.58	0.72	1.13	1.18
N6.2	*Feels unable to cope*	3.5	3.25	2.38	-0.04	0.25	0.22	0.44	0.55	0.84	1.74
O4.3	*Doesn't stick with only one way of doing things*	3.68	3.25	2.67	-0.05	0.21	0.25	0.49	0.46	0.9	1.02
A6.7	*Feels for the poor*	3.79	3.57	2.29	-0.05	0.22	0.34	0.58	0.5	0.82	1.4
C2.2	*Keeps possessions tidy*	3.61	3.32	2.38	-0.06	0.45	0.51	0.64	0.76	1.09	1.62
A3.3	*Is perceived as warm and uncalculating*	3.07	3.04	2.13	-0.06	0.26	0.31	0.57	0.68	1.01	1.4
E4.3	*Works quickly*	3.64	2.96	2.79	-0.06	0.31	0.43	0.53	0.55	1.05	0.81
N6.3	*Feels breaking down under stress*	3.57	3.04	2.88	-0.06	0.28	0.4	0.58	0.68	1.17	1.25
A1.6	*Defaults to trusting others*	3.29	2.86	2.54	-0.06	0.26	0.43	0.57	0.78	0.94	1.14
E1.8	*Cares for colleagues*	3.71	2.93	3	-0.06	0.22	0.28	0.45	0.67	0.84	1.44
C1.8	*Works efficiently*	3.5	3.21	2.25	-0.06	0.24	0.36	0.54	0.66	0.79	1.7

Traits and Adaptations

E3.3	*Emerges as leader*	3.71	3.18	2.75	-0.07	0.44	0.53	0.66	0.74	1.16	1.05
C4.3	*Gives up on self-improvements*	3.36	2.71	2.75	-0.07	0.32	0.35	0.54	0.6	1.07	1.36
E5.8	*Likes attending games*	3.64	2.68	3.25	-0.07	0.5	0.51	0.62	0.78	1.18	0.92
C1.1	*Is reasonable and practical*	3.57	3.21	2.29	-0.08	0.3	0.38	0.54	0.63	0.91	1.65
C2.4	*Puts things back where they belong*	3.57	3.11	2.46	-0.08	0.39	0.36	0.47	0.63	1.06	1.21
N2.1	*Is often riled by others*	3.43	2.68	2.83	-0.09	0.31	0.39	0.58	0.62	1.09	1.16
A3.4	*Tries to be respectful of others*	3.61	3.39	2.25	-0.09	0.22	0.27	0	0.43	0.76	1.71
A3.6	*Is generally liked*	3.32	2.89	2.46	-0.09	0.22	0.27	0.48	0.6	0.68	1.6
N1.7	*Has many fears*	3.21	2.86	2.38	-0.09	0.28	0.44	0.5	0.63	0.94	0.87
A4.1	*Prefers working together over rivalry*	3.86	3.04	3	-0.1	0.24	0.36	0.36	0.59	0.92	1.67
A2.3	*Could not deceive others*	3.18	2.71	2.46	-0.1	0.25	0.44	0.5	0.66	1.13	1.49
C1.4	*Comes prepared*	3.57	2.86	2.75	-0.1	0.22	0.35	0.6	0.5	0.95	0.93
C5.5	*Finishes what has been started*	3.82	3.14	2.67	-0.1	0.26	0.35	0.53	0.61	0.87	1.68
O3.1	*Values strong emotions*	3.29	2.61	2.71	-0.11	0.25	0.31	0.49	0.62	1.04	0.98
O1.3	*Has an active fantasy*	3.61	3	2.46	-0.11	0.31	0.47	0.55	0.72	1.09	1.03
E3.6	*Feels comfortable being a leader*	3.71	3.25	2.5	-0.11	0.33	0.47	0.51	0.62	1.06	0.85
E1.1	*Likes almost everyone*	3.29	3	2.29	-0.12	0.29	0.33	0.53	0.65	0.84	1.65
A4.6	*Finds it hard to fight back*	3.5	2.96	2.5	-0.12	0.27	0.46	0.56	0.74	1.11	1.24
E2.7	*Enjoys social events*	3.43	2.93	2.58	-0.12	0.37	0.43	0.58	0.68	0.96	1.19
N1.1	*Worries a lot*	3.5	3	2.46	-0.13	0.3	0.31	0.43	0.52	1.2	1.1

Traits and Adaptations

E1.4	*Is seen as approachable*	3.25	2.89	2.38	-0.13	0.29	0.44	0.54	0.74	1.01	1.35
N5.4	*Overeats favorite foods*	3.29	2.57	2.67	-0.13	0.36	0.48	0.6	0.68	1.08	0.88
C3.5	*Follows through on commitments*	3.93	3.11	2.63	-0.13	0.27	0.26	0.52	0.6	0.83	2.18
E6.5	*Is merry*	3.43	2.79	2.63	-0.13	0.4	0.41	0.53	0.48	1.04	1.21
C2.7	*Cleans excessively*	3.25	2.79	2.42	-0.13	0.41	0.45	0.5	0.61	1.15	0.82
A3.8	*Does anything to help*	3.75	3.32	2.42	-0.13	0.24	0.37	0.48	0.55	0.73	1.52
N6.4	*Gets dismayed in difficult situations*	3.93	3.5	2.25	-0.14	0.27	0.41	0.48	0.68	0.93	1.84
C4.5	*Wants to get ahead*	3.36	2.79	2.46	-0.14	0.3	0.46	0.43	0.55	1.12	1.24
N3.1	*Feels left behind*	3.29	2.79	2.29	-0.16	0.27	0.26	0.52	0.67	1.17	1.34
N5.8	*Cannot keep calm*	3.46	3.14	2.17	-0.16	0.29	0.35	0.42	0.59	1.04	1.18
A2.1	*Is honest*	3.25	2.64	2.38	-0.17	0.15	0.3	0.44	0.53	1.1	1.31
N5.3	*Finds it hard to resist temptations*	3.29	2.82	2.13	-0.17	0.27	0.24	0.39	0.4	1.05	1.18
N1.4	*Is often fidgety*	3.32	2.89	2.33	-0.17	0.35	0.39	0.56	0.74	1.07	1.32
O1.7	*Liked pretend play as a kid*	3.18	2.5	2.63	-0.17	0.23	0.42	0.58	0.53	0.99	0.81
E5.7	*Likes showy styles*	3.04	2.57	2.25	-0.17	0.46	0.37	0.52	0.64	1.08	0.66
C6.7	*Plans before travel*	4.04	2.79	3.04	-0.17	0.37	0.41	0.57	0.58	1.1	0.96
N1.8	*Has scary thoughts*	3.39	2.75	2.54	-0.18	0.26	0.44	0.56	0.64	1.13	0.98
C4.7	*Always strives for the best*	4.14	3.25	2.63	-0.19	0.31	0.43	0.56	0.6	1.01	1.19
E3.8	*Finds it easy to take lead*	3.61	2.79	2.63	-0.19	0.35	0.42	0.48	0.56	1.04	0.98
C4.1	*Is enthusiastic*	3.32	2.82	2.17	-0.2	0.32	0.34	0.49	0.53	0.98	1.51

Traits and Adaptations

E1.6	*Can easily approach strangers*	3.5	3.14	1.96	-0.21	0.35	0.39	0.56	0.65	1.05	1.66
C6.6	*Avoids hurried decisions*	3.46	2.79	2.38	-0.21	0.23	0.33	0.43	0.54	0.95	1.24
E2.8	*Likes crowded parties*	3.89	2.86	2.79	-0.21	0.5	0.56	0.61	0.82	1.17	0.96
C5.8	*Is very disciplined*	3.68	3.11	2.08	-0.21	0.33	0.48	0.53	0.62	0.96	1.45
O1.5	*Enjoys letting fantasies develop*	3.57	2.79	2.29	-0.22	0.28	0.41	0.58	0.75	1.13	1.04
E3.7	*Is the most talkative person in conversations*	3.68	2.71	2.5	-0.22	0.43	0.44	0.56	0.71	1.05	0.97
C4.6	*Is ambitious in everything*	3.96	3.04	2.42	-0.23	0.29	0.43	0.53	0.65	0.98	1.51
O4.8	*Is happy to try new routes*	3.71	2.86	2.42	-0.23	0.26	0.38	0.48	0.55	1.03	0.75
E6.3	*Sees the bright side*	3.46	2.96	2.17	-0.24	0.31	0.46	0.5	0.68	1.1	1.55
E5.3	*Does things just for a "kick"*	3.54	2.68	2.42	-0.25	0.41	0.48	0.59	0.64	1.14	0.89
C4.4	*Works hard towards his/her goals*	3.96	3.18	2.08	-0.25	0.33	0.33	0.59	0.64	0.93	1.77
C6.4	*Heeds consequences before doing something*	3.82	3.04	2.29	-0.25	0.23	0.34	0.55	0.66	1	1.51
E5.6	*Likes roller coasters*	3.68	2.29	2.88	-0.26	0.45	0.63	0.62	0.87	1.42	0.81
E3.1	*Is commanding*	3.75	3.14	2	-0.27	0.42	0.49	0.49	0.68	1.15	1.23
O4.5	*Is happy to change environment*	3.96	2.86	2.67	-0.27	0.27	0.4	0.53	0.55	1.03	0.69
C2.6	*Is somewhat pedantic and captious*	3.46	2.5	2.38	-0.27	0.28	0.43	0.45	0.61	1.04	1.11
O2.2	*Is entranced by music*	3.68	2.32	2.75	-0.28	0.4	0.5	0.67	0.8	1.15	1.32
E1.3	*Is friendly*	3.68	3.11	1.75	-0.28	0.27	0.47	0.5	0.76	0.8	1.76
E2.5	*Likes jobs that require working with others*	3.86	2.86	2.46	-0.28	0.3	0.44	0.53	0.65	1.1	0.79

N5.5	*Gives in to impulses*	3.75	2.82	2.25	-0.29	0.21	0.34	0.41	0.53	0.97	0.98
O5.5	*Enjoys puzzles*	3.61	2.29	2.75	-0.29	0.37	0.51	0.6	0.74	1.17	0.96
E1.2	*Enjoys gabbing*	3.39	2.5	2.13	-0.3	0.33	0.3	0.53	0.71	0.96	1.73
C2.5	*Is well organized*	3.32	2.39	2.21	-0.31	0.29	0.46	0.52	0.7	0.97	1.55
O3.8	*Has emotions provoked by a variety of things*	3.43	2.54	2.38	-0.31	0.31	0.43	0.57	0.59	1.1	0.87
E4.8	*Is always on the go*	3.61	2.68	2.08	-0.31	0.42	0.45	0.61	0.81	0.99	1.33
N3.3	*Is often sad*	3.61	2.61	2.25	-0.33	0.28	0.42	0.5	0.71	1.12	1.3
N2.4	*Is seen as a precarious person*	3.57	2.75	2.13	-0.33	0.24	0.26	0.4	0.38	1.01	1.2
O5.7	*Desires to understand things*	3.75	2.79	2.04	-0.34	0.36	0.49	0.53	0.7	0.9	1.51
O5.1	*Enjoys thinking theoretically*	3.82	2.64	2.33	-0.35	0.32	0.54	0.57	0.73	1.17	0.99
E2.2	*Likes to be surrounded by people*	3.68	2.54	2.38	-0.35	0.43	0.53	0.62	0.74	1.09	1.26
N1.3	*Feels anxiety*	3.64	2.93	1.83	-0.35	0.32	0.46	0.54	0.71	1.09	1.08
E2.1	*Feels comfortable in crowds*	3.68	2.61	2.21	-0.36	0.4	0.44	0.59	0.71	1.24	0.98
C2.3	*Is meticulous*	3.54	2.57	2	-0.36	0.29	0.48	0.54	0.58	1.04	1.07
O1.8	*Can let mind wander*	3.68	2.39	2.42	-0.36	0.16	0.32	0.42	0.36	1	0.7
O2.4	*Is interested in patterns*	3.93	2.68	2.29	-0.37	0.44	0.44	0.6	0.75	1.1	1.13
E4.5	*Is no less energetic than others*	3.18	2.43	1.88	-0.37	0.34	0.37	0.56	0.66	1.06	0.82
E2.3	*Prefers company*	3.89	3.18	1.79	-0.38	0.36	0.46	0.54	0.64	1.1	0.98
N2.6	*Is easy to infuriate*	3.96	2.79	2.21	-0.38	0.35	0.45	0.51	0.7	1.08	1.41
C5.4	*Can make him/herself do what needs doing*	3.71	2.46	2.21	-0.39	0.28	0.42	0.54	0.62	1.08	1.3

A6.2	*Is not hard-nosed*	3.86	2.96	1.83	-0.4	0.29	0.36	0.54	0.57	1.03	1.25
C6.2	*Avoids hasty decisions*	4.07	3.11	1.92	-0.4	0.27	0.35	0.53	0.45	0.88	1.55
N1.2	*Is easily frightened*	3.57	2.96	1.54	-0.41	0.35	0.43	0.6	0.74	1.13	1.05
E6.2	*Is sometimes overwhelmed by joy*	3.79	2.68	2.13	-0.41	0.3	0.38	0.2	0.59	0.86	1.58
E2.4	*Starts longing for company when alone*	3.68	2.64	2.04	-0.41	0.3	0.41	0.5	0.48	1.06	1.34
E1.7	*Is emotionally attached to friends*	4.11	2.93	2.33	-0.41	0.32	0.41	0.54	0.73	0.91	1.52
O4.1	*Can easily change habits*	3.5	2.5	2	-0.42	0.23	0.45	0.57	0.6	1.03	1.07
O3.3	*Minds his/her feelings*	3.96	2.79	2.17	-0.42	0.23	0.29	0.48	0.42	0.79	0.98
N2.8	*Is easy to frustrate*	3.71	2.57	2.04	-0.43	0.27	0.4	0.57	0.59	1.07	1.4
N6.8	*Is emotionally unstable*	3.82	2.61	2	-0.43	0.34	0.32	0.34	0.78	0.95	1.56
E4.2	*Acts strenuously*	3.93	2.71	2	-0.44	0.33	0.39	0.48	0.57	0.82	1.6
N5.2	*Cannot resist cravings*	3.79	2.68	1.96	-0.44	0.24	0.44	0.55	0.68	1.08	1.06
N6.5	*Finds it hard to decide*	3.57	2.36	2.08	-0.44	0.31	0.42	0.54	0.7	1.11	1.07
N2.3	*Is short-tempered*	3.61	2.82	1.63	-0.45	0.4	0.46	0.58	0.7	1.14	1.56
O5.3	*Likes mental challenges*	4	2.89	1.92	-0.45	0.35	0.44	0.61	0.62	1.09	1.18
N5.7	*Does things offhand and later regrets*	3.86	2.61	2.21	-0.45	0.22	0.37	0.47	0.66	1.01	0.98
C6.3	*Sometimes acts without thinking*	3.54	2.36	2.04	-0.46	0.24	0.4	0.41	0.66	1.09	1.11
E6.4	*Is sometimes extremely happy*	3.57	2.43	2.08	-0.46	0.32	0.51	0.5	0.74	0.91	1.61
N1.6	*Keeps thinking of what might go astray*	4.04	2.79	1.96	-0.46	0.32	0.39	0.56	0.72	1.08	0.98
O3.2	*Experiences strong feelings*	4	2.57	2.25	-0.49	0.27	0.4	0.49	0.59	1.04	1.09

C6.5	*Acts impromptu*	3.79	2.46	2.04	-0.49	0.3	0.44	0.52	0.69	1.01	0.93
N2.2	*Is agitated*	3.82	2.71	1.63	-0.51	0.32	0.45	0.45	0.65	1	1.51
O3.7	*Can easily feel other's emotions*	3.89	2.68	1.83	-0.51	0.22	0.4	0	0.59	0.88	1.53
E5.1	*Craves for excitement*	4.04	2.36	2.21	-0.52	0.34	0.43	0.59	0.69	1.09	0.81
E6.6	*Is exuberant*	3.86	2.5	1.92	-0.52	0.37	0.56	0.58	0.78	0.93	1.75
E5.5	*Wants action*	4	2.75	1.88	-0.53	0.4	0.44	0.57	0.7	1.05	1.04
E6.8	*Is easy to make laugh*	3.93	2.46	2.08	-0.54	0.34	0.42	0.45	0.71	0.91	1.52
O1.1	*Has bustling imagination*	4	2.68	1.54	-0.6	0.31	0.4	0.55	0.75	1.05	1.23
E1.5	*Loves talking to people*	4.11	2.5	1.75	-0.64	0.39	0.41	0.58	0.8	0.91	1.72
E4.4	*Is full of energy*	3.68	1.89	1.83	-0.66	0.34	0.41	0.6	0.76	1.01	1.32
A4.7	*Is persistent and unyielding*	4.04	2.43	1.5	-0.72	0.31	0.45	0.49	0.75	1.13	1.3
O3.5	*Has many emotions*	4.18	1.82	1.63	-0.88	0.22	0.42	0.47	0.58	0.89	1.19

Note. Code refers to an item's notation in the NEO-PI-R and are interpretable as Big Five domain (N = Neuroticism, E = Extraversion, O = Openness to Experience, A = Agreeableness, C = Conscientiousness), facet (1-6 for each domain), and item (1-8 for each facet), respectively. T_c, A_c, V, and A_O refer to rating conditions *Combined Trait, Combined Adaptation, Bipolar,* and *Overall Adaptation-ness*. A_O scores are the result of averaging across means of standardized item scores within the original 5 rating conditions. r_{CA} = cross-rater agreement. r_{RO} = rank-order stability. h^2 = heritability. r_{TT} = 7-10-day re-test reliability. *Var* = item standard deviations. *Eval* = overall evaluativeness (absolute value of social desirability).

Table A2. *Descriptive statistics and Spearman's correlations for all variables.*

Variable	*M*	*SD*	Agreement	Stability	Heritability	Variance	Evaluativeness	Re-test
Agreement	.30	.07						
Stability	.40	.07	.61***					
Heritability	.52	.10	.49***	.51***				
Variance	1.03	.13	.52***	.54***	.43***			
Evaluativeness	1.25	.31	-.21**	-.27***	-.19*	-.48***		
Re-test Reliability	.64	.09	.57***	.56***	.49***	.38***	.01	
Trait	3.38	.40	0.11	0.11	-0.06	-0.18**	0.10	0.01
Adaptation	3.09	.44	-0.29***	-0.23***	-0.11	-0.19**	0.27***	-0.12
Bipolar	2.68	.57	-0.09	-0.08	0.04	0.15*	-0.16*	-0.04
Adaptation (Overall)	0	.35	-0.20**	-0.17**	-0.01	0.03	0.02	-0.06

Note. M = mean. SD = standard deviation. "Trait," "Adaptation," "Bipolar," and "Adaptation (Overall)" refer to T_c A_c, V, and A_O ratings. "Variance" = item standard deviations. "Re-test" = 7-10-day re-test reliability. Intercorrelations among BT-CA rating conditions are not included as they are previously recorded in Table 3. * $p < .05$. ** $p < .01$. *** $p < .001$.